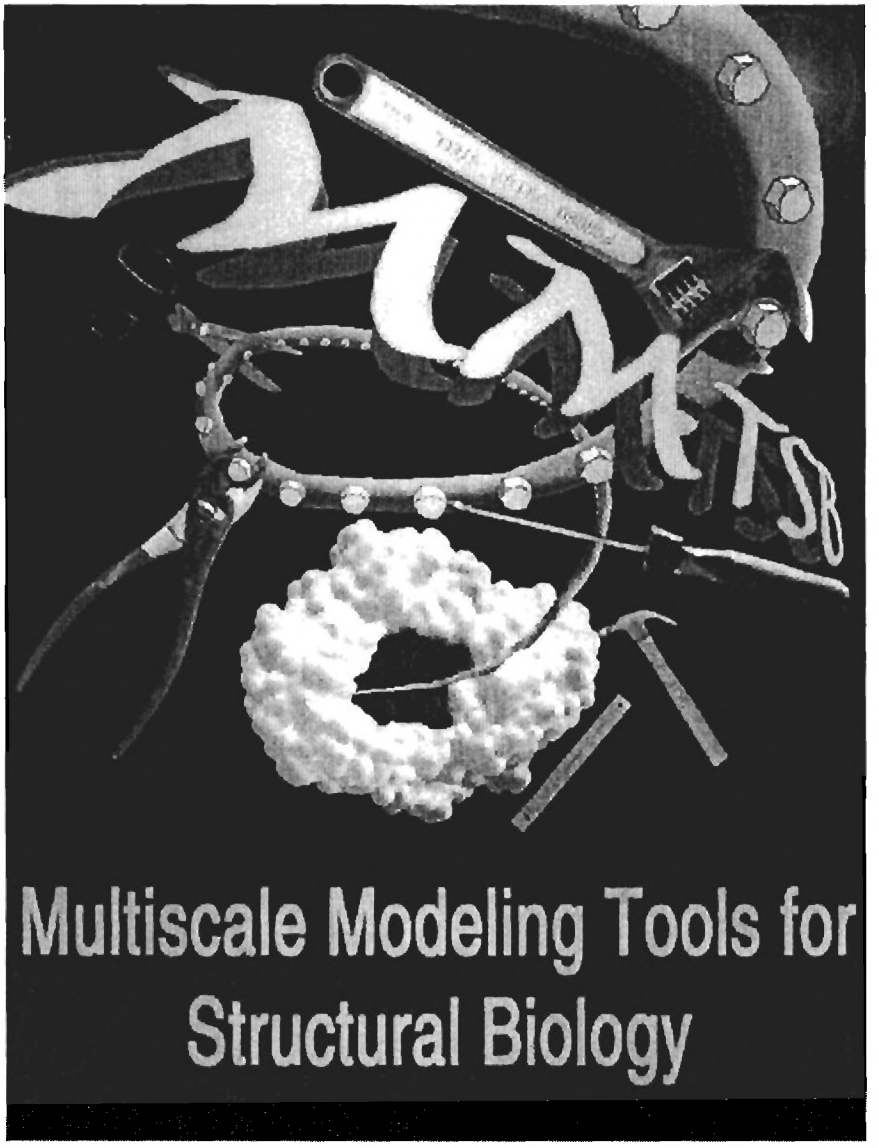
G-32-6EW
#1

Center for the development of Multi-scale Modeling Tools for Structural Biology

Annual Progress Report
July 29, 2005



http://mmtsb.scripps.edu

NATIONAL INSTITUTES OF HEALTH
DEPARTMENT OF HEALTH AND HUMAN SERVICES

BIOMEDICAL TECHNOLOGY RESOURCE GRANTS
DIVISION OF BIOMEDICAL TECHNOLOGY
NATIONAL CENTER FOR RESEARCH RESOURCES

5P41RR012255-08

MULTISCALE MODELING TOOLS FOR STRUCTURAL BIOLOGY

Final

SCRIPPS RESEARCH INSTITUTE

ANNUAL PROGRESS REPORT

Reporting From: 08/31/2004
Reporting To: 08/31/2005

0.000% AIDS Related

Signature Date
CHARLES BROOKS, BS, PHD
PROFESSOR
Phone: 858-784-8035
Fax: 858-784-8055
e-mail: brooks@scripps.edu

Patent or Copyright was not awarded this grant year.

TABLE OF CONTENTS

PROTECTION AGAINST RESEARCH RISKS

No	Will human subjects or human data be used over the next year?
N/A	Is IRB approval current for all participating research sites?
N/A	Has human subjects training been completed by all key personnel?
N/A	Has a data and safety monitoring plan been put in place?
Yes	Have the gender/minority requirements for recruitment/retention been met
No	Will vertebrate animals be used over the next year?
N/A	Have IACUC approvals been obtained for all participating research sites?

PERSONNEL ROSTER

Name, Degree	Department, Non-Host Institution Information
AN, CHAHM, BS	MOLECULAR BIOLOGY
ANDERSON, DAVID, BS, PHD	SPACE SCIENCE LABORATORY UNIVERSITY OF CALIFORNIA, BERKELEY CA USA
BAJAJ, CHANDRAJIT, BS, PHD	COMPUTER SCIENCE UNIVERSITY OF TEXAS, AUSTIN TX USA
BAKER, TIM, BS, PHD	BIOLOGY UNIVERSITY OF CALIFORNIA, SAN DIEGO CA USA
BERMAN, HELEN	CHEMISTRY RUTGERS UNIVERSITY NJ USA
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
CARRAGHER, BRIDGET, BS, PHD	CELL BIOLOGY
CASE, DAVID, AM, PHD, BS	MOLECULAR BIOLOGY
CHEN, JIANHAN, BS, PHD	MOLECULAR BIOLOGY
CUI, QIZHI, BS, PHD	MOLECULAR BIOLOGY
DYSON, H JANE	MOLECULAR BIOLOGY
FALKE, SCOTT, BS, PHD	BIOLOGY AND MOLECULAR BIOLOGY UNIVERSITY OF KANSAS MEDICAL CENTER KS USA
FEIG, MICHAEL, BS, PHD	BIOCHEMISTRY AND CHEMISTRY MICHIGAN STATE UNIVERSITY MI USA
FISHER, MARK, BS, PHD	BIOLOGY AND MOLECULAR BIOLOGY UNIVERSITY OF KANSAS MEDICAL CENTER KS USA
FRANK, JOACHIM, BS, PHD	BIOMEDICAL SCIENCES SUNY, ALBANY AND HOWARD HUGHS MEDICAL INSTITUTE NY USA
GAO, JIALI, BS, PHD	CHEMISTRY UNIVERSITY OF MINNESOTA MN USA
GOGOL, EDWARD, BS, PHD	BIOLOGICAL SCIENCES UNIVERSITY OF MISSOURI - KANSAS CITY KS USA
HARVEY, STEPHEN, BS, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY GA USA

Name, Degree	Department, Non-Host Institution Information
IM, WONPIL, BS, PHD	MOLECULAR BIOLOGY
JOHNSON, JOHN E, BS, PHD	MOLECULAR BIOLOGY
KONECNY, ROBERT, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD CA USA
LIU, JUN	BIOMEDICAL SCIENCES FLORIDA STATE UNIVERSITY FL USA
LOCKER, REBECCA C, BS, PHD	BIOLOGICAL SCIENCES GEORGIA TECHNOLOGICAL INSTITUTE GA USA
LUI, H, BS, PHD	MOLECULAR BIOLOGY
MACKERELL, ALEXANDER	PHARMACEUTICAL CHEMISTRY UNIVERSITY OF MARYLAND, SCHOOL OF PHARMACY MD USA
MCCAMMON, JAMES, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD CA USA
MITRA, ALOK, BS, PHD	BIOLOGICAL SCIENCES UNIVERSITY OF AUCKLAND NEW ZEALAND
MIYASHITA, OSAMU, BS, PHD	PHYSICS UCSD CA USA
POTTER, CLINT, BS, PHD	CELL BIOLOGY
QU, C, BS, PHD	MOLECULAR BIOLOGY
REDDY, VIJAY S, BS, PHD	MOLECULAR BIOLOGY
SHEPHERD, CRAIG M, BS, PHD	MOLECULAR BIOLOGY
SKOLNICK, JEFFREY, BA, PHD	CENTER FOR EXCELLENCE IN BIOIN SUNY, BUFFALO NY USA
SMITH, THOMAS, BS, PHD	DONALD DANFORTH PLANT SCIENCE CENTER MO USA
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
TAN, ROBERT K Z, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY GA USA
TAUFER, MICHELA, BS, PHD	MOLECULAR BIOLOGY

Name, Degree	Department, Non-Host Institution Information
TAYLOR, KENNETH, BS, PHD	BIOMEDICAL SCIENCES FLORIDA STATE UNIVERSITY FL USA
TRYLSKA, JOANNA, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD CA USA
VALLE, MIKEL, BS, PHD	BIOMEDICAL SCIENCES SUNY, ALBANY AND HOWARD HUGHES MEDICAL INSTITUTE NY USA
WHITE, STEPHEN S., BS, PHD	PHYSIOLOGY AND BIOPHYSICS UNIVERSITY OF CALIORNIA, IRVINE CA USA
WILLIAMSON, JAMES, BS, PHD	MOLECULAR BIOLOGY

SUBPROJECT DESCRIPTIONS

COLLABORATIVE RESEARCH

ASSEMBLY ENERGETICS OF THE 30S RIBOSOMAL SUBUNIT (0023)

BTR Unit: Collaborative Research
%BTR $: 0.500%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
MCCAMMON, JAMES, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD, CA USA
TRYLSKA, JOANNA, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD, CA USA

Subproject Description

To explore the relationship between the assembly of the 30S ribosomal subunit and interactions among the constituent components, 16S RNA and proteins, relative binding free energies of the T. thermophilus 30S proteins to the 16S RNA were studied based on an implicit solvent model of electrostatic, nonpolar, and entropic contributions. The late binding proteins in our assembly map were found not to bind to the naked 16S RNA. The 5Å domain early kinetic class proteins, on average, carry the highest positive charge, get buried the most upon binding to 16S RNA, and show the most favorable binding. Some proteins (S10/S14, S6/S18, S13/S19) have more stabilizing interactions while binding as dimers. Our computed assembly map resembles that of E. coli; however, the central domain path is more similar to that of A. aeolicus, a hyperthermophilic bacteria.

ELECTROSTATIC CONTROL OF VIRUS CAPSID SWELLING IN CCMV (0024)

BTR Unit: Collaborative Research
%BTR $: 0.500%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
KONECNY, ROBERT, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD, CA USA
MCCAMMON, JAMES, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD, CA USA
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
TRYLSKA, JOANNA, BS, PHD	CHEMISTRY AND BIOCHEMISTRY UCSD, CA USA

Subproject Description

Electrostatic properties of cowpea chlorotic mottle virus (CCMV) and cucumber mosaic virus (CMV) were investigated using numerical solutions to Poisson-Boltzmann equation. Experimentally it has been shown that CCMV particles swell in the absence of divalent cations when the pH is raised from 5 to 7. CMV, although structurally homologous, does not undergo this transition. Analysis of the calculated electrostatic potential confirms that a strong electrostatic repulsion at the calcium binding sites in the CCMV capsid is most likely the driving force for the capsid swelling process. The binding interaction between encapsulated genome material (RNA) inside of the capsid and the inner capsid shell is weakened during the swelling transition. This probably aids in the RNA release process but it is unlikely that RNA is released through capsid openings due to unfavorable electrostatic interaction between RNA and capsid inner shell residues at these openings. Differences in the calcium binding energies to the CCMV native and swollen structures were also calculated. Both forms show favorable calcium binding energies, with the native form being at least three times more stable. The results suggest that Ca2+ ions can also bind to the swollen form and thus may induce the capsid contraction.

NMR REFINEMENT (0028)

BTR Unit: Collaborative Research
%BTR $: 2.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
CHEN, JIANHAN, BS, PHD	MOLECULAR BIOLOGY
DYSON, H JANE,	MOLECULAR BIOLOGY
IM, WONPIL, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

Electrostatic interactions are often oversimplified or ignored in the energy functions for NMR structure calculations, because it is difficult to evaluate them reliably without proper description of the dielectric screening by solvent. In light of recent improvements in implicit solvent models, it was showed that simulated annealing refinement in a GB implicit solvent could lead to noticeable improvement in the final protein NMR structures in terms of the backbone dihedral angle distributions and hydrogen bond patterns. However, the impact of implicit solvent is rather small when a sufficient number of experimental restraints exist (such as in the final stage of NMR structure determination).

Subproject Progress

It was recently demonstrated that replica exchange molecular dynamics (REX-MD) refinement in a GB implicit solvent model could significantly improve the quality of structures and the radius of convergence when the experimental data is limited. For example, while conventional structure calculations using an initial set of sparse NOE restraints were unable to identify a unique topology for a protein domain, high-quality native-like initial folds were generated through REX-MD refinement of the initial structures with a GB implicit solvent. These models could be then used to make further assignments of ambiguous NOEs and speed up the structure determination process.

RIBOSOME STRUCTURAL STUDIES (0013)

BTR Unit: Collaborative Research
%BTR $: 1.000%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
FRANK, JOACHIM, BS, PHD	BIOMEDICAL SCIENCES SUNY, ALBANY AND HOWARD HUGHS MEDICAL INSTITUTE, NY USA

Subproject Description

The proper functioning of any living cell requires that approximately half of all proteins synthesized in the cytosol must be translocated across (in the case of soluble proteins) or inserted into (in the case of membrane proteins) a cell membrane. Protein translocation at the membrane occurs through a proteinaceous channel, termed the translocon or translocase. The core of the translocon, the protein-conducting channel (PCC), is heterotrimeric integral membrane protein complex (SecY in eubacteria/archaea, Sec61 in eukaryotes) composed of the following: the gamma-subunit (SecY in eubacteria/archaea, Sec61? in mammals), the gamma-subunit (SecG in eubacteria, Sec? in archaea, Sec61? in mammals) and the ?-subunit (SecE in eubacteria/archaea, Sec61? in mammals).

Cryo-EM reconstruction, performed by J. Frank and his collaborators, of the E. coli PCC, SecYEG, complexed with the ribosome and a signal sequence containing nascent chain, shows mRNA, three tRNAs, the nascent chain and detailed features of both an open, translocating PCC and a second, closed PCC bound to mRNA hairpins. Th translocating PCC forms connections with ribosomal RNA hairpins on two sides and ribosomal proteins at the back, leaving a frontal opening. Normal mode flexible fitting was used to interpret the cryo-EM reconstruction in term of high-resolution structure. NMFF of the archaeal SecYE? structure into the PCC EM densities favors a front-to-front arrangement of two SecYEG complexes in the PCC, and supports channel formation by the opening of two linked SecY halves during polypeptide translocation. Based on this observation in the open PCC of two segregated channels with different degrees of access to bulk lipid, a model for co-translational translocation was proposed

STRUCTURES & MECHANICS OF MYOSIN (0014)

BTR Unit: Collaborative Research
%BTR $: 2.000%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
LIU, JUN,	BIOMEDICAL SCIENCES FLORIDA STATE UNIVERSITY, FL USA
TAYLOR, KENNETH, BS, PHD	BIOMEDICAL SCIENCES FLORIDA STATE UNIVERSITY, FL USA

Subproject Description

A combination of experimental structural data, homology modelling and elastic network normal mode analysis have been used to explore how coupled motions between the two myosin heads and the dimerization domain (S2) in smooth muscle myosin II determine the domain movements required to achieve the inhibited state of this ATP dependent molecular motor. These physical models rationalize the empirical requirement for at least two heptads of noncoiled alpha-helix at the junction between the myosin heads and S2 and the dependence of regulation on S2 length. The results correlate well with biochemical data regarding altered conformational dependent solubility and stability. Structural models of the conformational transition between putative active states and the inhibited state show that torsional flexibility of the S2 ?-helices is a key mechanical requirement for myosin II regulation. The results may have implications for function in other dimeric molecular motors.

FLEXIBLE REFINEMENT OF ATOMIC LEVEL STRUCTURE INTO THE GROEL-SUBSTRATE COMPLEX (0021)

BTR Unit: Collaborative Research
%BTR $: 0.500%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
FALKE, SCOTT, BS, PHD	BIOLOGY AND MOLECULAR BIOLOGY UNIVERSITY OF KANSAS MEDICAL CENTER, KS USA
FISHER, MARK, BS, PHD	BIOLOGY AND MOLECULAR BIOLOGY UNIVERSITY OF KANSAS MEDICAL CENTER, KS USA
GOGOL, EDWARD, BS, PHD	BIOLOGICAL SCIENCES UNIVERSITY OF MISSOURI - KANSAS CITY, KS USA

Subproject Description

The 13Å resolution structures of GroEL bound to a single monomer of the protein substrate glutamine synthetase (GSm), as well as that of unliganded GroEL have been determined from a heterogeneous image population using cryo-electron microscopy (cryo-EM) coupled with single particle image classification and reconstruction techniques. Structural data from cryo-EM maps and dynamic modeling were combined, taking advantage of the known X-ray crystallographic structure and normal mode flexible fitting (NMFF) analysis, to describe the changes that occur in GroEL structure induced by GSm binding. The NMFF analysis reveals that the molecular movements induced by GSm binding propagate throughout the GroEL structure. The modeled molecular motions show that some domains undergo en-bloc movements while others show more complex independent internal movements. Interestingly, the substrate-bound apical domains of both the cis (GS-bound ring) and trans (opposite substrate free ring) show counter-clockwise rotations, in the same direction (though not as dramatic) as those documented for the ATP-GroEL induced structure changes. The structural changes from the allosteric substrate protein-induced negative cooperativity between the GroEL rings involves upward concerted movements of both cis and trans equatorial domains toward the GSm bound ring while the interring distances between the heptamer contact residues are maintained. Furthermore, the NMFF analysis also identifies the secondary structural elements that are involved in the observed ~5 Å reduction in the diameter of the cavity opening in the unbound trans ring. Understanding the molecular basis of these substrate protein-induced structural changes across the heptamer rings provides insight into the origins of the allosteric negative cooperative effects that are transmitted over long distances (~140 Å).

RED CLOVER NECROTIC MOSAIC VIRUS (0025)

BTR Unit: Collaborative Research
%BTR $: 0.500%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BAKER, TIM, BS, PHD	BIOLOGY UNIVERSITY OF CALIFORNIA, SAN DIEGO, CA USA
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

Red clover necrotic mosaic virus is a small, icosahedral plant virus (Tombusviridae family) with a single-stranded RNA genome. RCNMV virions were analyzed by cryo-electron microscopy (cryoEM) and three-dimensional reconstruction at 8 Å resolution. The virion capsids have prominent protrusions; inside the capsid there is an inner cage formed by complexes of genomic RNA and amino terminal portion of the coat protein. Separate reconstructions were obtained at 12 Å and 16 Å resolution, respectively, for virion samples that had Ca2+ alone or both Mg2+ and Ca2+ divalent cations selectively extracted. Virions treated in this way exhibited dramatic conformational changes, as obtained from NMFF, in both capsid and inner cage, though the particle diameters remained unchanged. Removal of Ca2+ ions alone appears to trigger movement and rotation of the S and P domains at the quasi 3-fold axis of symmetry. Additional removal of Mg2+ led to the formation of 11-13 Å diameter channels that extend through the virion shell at the quasi 3-fold axis of symmetry and noticeably alter the inner cage conformation. Despite their relatively small size, these channels are sufficiently large to allow RNA to pass through the protein capsid shell. Structural and infectivity studies suggest the divalent cation induced dynamics may be involved in viral reorganization, which renders the viral RNA accessible for translation early in infection.

DISSEMINATION & TRAINING

SOFTWARE DVMT, DOCUMENT & DISSEM WEB PAGES OF INFO ON MULTI SCALE MODELING TOOL (0011)

BTR Unit: Dissemination and Training
%BTR $: 5.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
FEIG, MICHAEL, BS, PHD	BIOCHEMISTRY AND CHEMISTRY MICHIGAN STATE UNIVERSITY, MI USA

Subproject Description

The MMTSB web page (http://mmtsb.scripps.edu) is maintained to provide web services to the scientific community, distribute and support software tools developed within the resource, offer general information about the MMTSB resource and its projects, and provide selected links to related resources. Already available web services include the very successful and established Virus Particle Explorer (VIPER) as well as some utility functions for nucleic acid and protein modeling applications, in particular for interconversion between reduced lattice models and all-atom representation and the Go-model server.

The MMTSB web site itself is the distribution site for the MMTSB Tool Set that is being developed for multiscale protein modeling and structure prediction application. It is a collection of modules, mostly written in perl to ensure portability and allow for easy modifications and extensions. This tool set provides interfaces to CHARMM, Amber, and MONSSTER specifically geared towards protein modeling applications and implements new methods developed within the resources for model interconversion and advanced sampling techniques. The design of the tool set is highly modular offering easy-to-use application-oriented utilities as well as flexible programming interfaces for more complex customized applications. Although still under development the current version of the MMTSB tool set is quite complete with respect to the available methods. Recently we have expanded its functionality towards a more comprehensive support of the Amber molecular modeling package.

Subproject Progress

More recently we have released the Predictor@home web site for volunteer distributed computing protein structure prediction. We are preparing additional web services for protein structure evaluation, structure refinement, and limited protein structure prediction with a focus on loop modeling as the development and testing of the necessary modeling tools nears completion and additional computational resources needed for this kind of service are coming online. Several software packages for modeling applications are supported by the resource with links on the MMTSB web site to the corresponding distribution and documentation sites. They include CHARMM and Amber for general modeling applications, MONSSTER for lattice-based protein modeling, NAB, Yammp and YUP for nucleic acid modeling.

In addition, the software package (NMFF) for the flexible refinement of high-resolution X-ray structures into low resolution data from cryo-electron microscopy is to be released for distribution and service related refinement projects.

REWRITING YAMMP (0018)

BTR Unit: Dissemination and Training
%BTR $: 7.000%

Investigator	Department/Non-Host Information
HARVEY, STEPHEN, BS, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY, GA USA
CASE, DAVID, AM, BS, PHD	MOLECULAR BIOLOGY
CUI, QIZHI, BS, PHD	MOLECULAR BIOLOGY
TAN, ROBERT K Z, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY, GA USA

Subproject Description

The rewriting of Yammp into YUP (Yammp Under Python) is complete, and YUP has been released. It is available from our website (http://rumour.biology.gatech.edu/YammpWeb/).The rewrite achieved two long-standing goals: (1) integration of a scripting language and (2) the ability to assemble the force field for all types of models and especially non-standard ones. In the new version of the program, energy and force calculations form the computational core, and these functions remain written in the C programming language and compiled into a number of shared libraries of native machine code. Molecular mechanics methods such as energy minimization and molecular dynamics (MD) are implemented as Python programs that call the shared code.

Subproject Progress

First, we developed a prototype graphical user interface for Yammp (YupSee). It has two major components, one for setting up, carrying out, and analyzing simulations, and the other for displaying the resulting models. We have decided that the further development of YupSee will require an effort and resources that are beyond the scope of this grant, so we are seeking separate funding for this. (An R01 proposal was submitted on June 24, 2005 for this purpose, in response to PAR-03-106.)

Second, we have begun major efforts at Yammp dissemination. A series of five tutorials has been developed and posted online (http://rumour.biology.gatech.edu/YammpWeb/). The goal is to develop these to the point where they are stand-alone exercises, so that new users can download Yammp and the tutorials from the web, work their way through the tutorials, and begin using Yammp without assistance from anyone at Georgia Tech. Before this can be done, however, we are running a series of workshops to train new users, and to gain feedback for revising the tutorials. The first YUP training workshop was run at Georgia Tech from June 28 ? July 1, 2005, and the second was run July 5-8. Each of these had five participants. As a result of user feedback, we have expanded the first tutorial (which teaches the use of Python) and are requiring future trainees to complete this before coming to the workshops. This enables us to shorten the workshop itself to three days. A third workshop is scheduled for August 16-18, and four people are currently enrolled.

SERVICES

SIMPLIFIED MODELS FOR PROTEIN FOLDING KINETICS & THERMODYNAMICS & MECHANISM (0009)

BTR Unit: Services
%BTR $: 1.500%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

This project is directed toward the development of rapid low-resolution models for protein folding studies exploring both folding kinetics and thermodynamics. Progress has been made in addressing questions relating to protein folding kinetics and mechanisms using simple models. A procedure has been developed to distill an all-atom protein structure into a set of potentials, which then are applied to a simple protein representation. These simplified protein models have been characterized in detail for over 20 proteins to date, and have proved particularly helpful in identifying the origins of experimentally observed qualitative differences in the folding of proteins of shared topology. These models have also been shown to offer considerable advantage over a related class of Ising-like models that depend only on the protein topology. The algorithm for generating such potentials has been implemented as a web-based tool (http://mmtsb.scripps.edu/webservices/gomodel.html), offering users a set of potentials for studying their protein of interest.

Subproject Progress

The web service of our Go modeler continues to be improved and is being used by an increasing number of scientists around the world.

FLEXIBLE REFINEMENT OF ATOMIC LEVEL STRUCTURE INTO THE LF-PA63H COMPLEX (0022)

BTR Unit: Services
%BTR $: 1.500%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
MITRA, ALOK, BS, PHD	BIOLOGICAL SCIENCES UNIVERSITY OF AUCKLAND, NEW ZEALAND

Subproject Description

Anthrax is caused by one of the two virulence factors produced by the bacteria Bacillus anthracis. This factor is a cocktail of three monomeric proteins, lethal factor (LF, 90kD), edema factor (EF, 89kD) and the protective antigen (PA, 83kD).

Understanding the molecular recognition processes involved in the formation of the toxic Antrahx complex and the eventual delivery of the lethal and edema factors is critical to gaining insight into the pathogenesis of anthrax, and eventually to rational design of agents to block toxin action. Toward this end, we had directly visualized the PA63h.LF complex from analysis of images of specimens suspended in vitrified buffer by cryo-electron microscopy (Ren et al., 2004). Comparison of the averaged 3-dimensional (3-D) views of the liganded and unliganded complex showed that the PA63h molecule binds to one LF molecule, which is localized to the non-membrane-interacting ?top? face of the heptamer. The observed asymmetric binding of the ligand involves interaction with successive (at least 3) PA63 monomers and causes partial unravelling of the heptamer enlarging the lumen volume (Ren et al., 2004). This structural perturbation could be utilized to allow the ligand to ease into the vestibular lumen prior to its eventual translocation through the pore in the membrane.

Subproject Progress

The docking of one molecule of LF and the seven PA63 monomers in the density map of the complex was visually carried out, using the program O, based on the X-ray structures of PA63 and PA63h (Petosa et al., 1997) and LF (Pannifer et al., 2001) and using knowledge of the interacting regions of the two-polypeptide chains, as determined by site-directed mutagenesis studies. These, for PA63 have been identified (Cunningham et al. 2002) as well as for LF (Lacy et al. 2002). Based on the volume and shape of the density attributed to the LF in the cryo EM map generated by Ren et al. (2004), it was clear that LF has very likely undergone conformational rearrangement upon binding to the PA63 heptamer. In this study we have reanalyzed the model of the liganded complex using the normal mode flexible fitting (NMFF) procedure. In this technique, a linear combination of low-frequency normal modes, that are known to well represent conformational change in a molecule, is used in an iterative manner to deform the modeled structure optimally to conform to the low to medium resolution electron density map. Using NMFF and the previously determined 18Å resolution density map of the PA63h:LF complex (Ren et al., 2004), a model for the ligand bound structure is constructed. The details of the heptamer structure, the observed conformational change of the LF molecule, and the nature of the ligand binding interface is discussed vis-à-vis their possible implication in the biological activity of anthrax toxin.

TECHNOLOGICAL RESEARCH & DEVELOPMENT

PROTEIN MODELING & STRUCTURAL GENOMICS (0010)

BTR Unit: Technological Research and Development
%BTR $: 4.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
CASE, DAVID, AM, BS, PHD	MOLECULAR BIOLOGY
FEIG, MICHAEL, BS, PHD	BIOCHEMISTRY AND CHEMISTRY MICHIGAN STATE UNIVERSITY, MI USA

Subproject Description

The goal of the protein modeling project of the MMTSB resource aims at developing such a multi-scale modeling approach to protein structure prediction, fragment and loop modeling, as well as structure refinement of approximatc structures derived by other methods. For lattice-based low-resolution modeling the MONSSTER program by Kolinski and Skolnick is used where proteins are represented with a single particle for each residue located at the side chain center and simulated with a Monte Carlo scheme on a cubic lattice. All-atom modeling is performed using CHARMM with an implicit solvent model based on a Generalized Born approximation. This allows rapid energy evaluations for all-atom structures and offers the possibility of fast all-atom simulations for further structure refinement. Efforts during the past year have focused on exploring the accuracy and speed of generalized Born models, developed under support from the resource, in the context of protein scoring calculations. Additional efforts have aimed to ¿harden¿ the distributed MMTSB Tool Set for protein modeling and structure prediction.

Subproject Progress

We continue to develop and release new versions of the MMTSB Tool Set. During the past year we have added further user suppor through the CHARMM forum at http://www.charmm.org. Additioally, we deveoped and ran a new workshop based on the MMTSB Tool Set and associated theory and computational modeling last August in San Diego. The workshop, which served about 45 students, was over subscribed and well received. Detailed tutorials and lecture notes from the workshop can be found at http://ctbp.ucsd.edu/workshopinfo_2004.html. The workshop was co-organized with the Center for Theoretical Biological Physics.

IMPLICIT SOLVATION MODELS FOR PROTEINS, NUCLEIC ACIDS & MEMBRANES (0012)

BTR Unit: Technological Research and Development

%BTR $: 5.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
CHEN, JIANHAN, BS, PHD	MOLECULAR BIOLOGY
IM, WONPIL, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

This core research project is aimed at the development of implicit solvent models for representation of aqueous and membrane environments for the purposes of exploring the folding and modeling of integral membrane proteins and peptides. The focus has been in the exploration of generalized Born implicit solvent/implicit membrane models and extensions of the treatment of long-range dispersion interactions in these systems. Applications to the folding of proteins and peptides in aqueous solution, the refinement of structure in the presence of NMR derived restraints, the folding and structure determination of membrane bound peptides and proteins and the development of homology models

Subproject Progress

Reformulating the calculation of the self-electrostatic solvation energy in GBSW (generalized Born model a simple smoothed switching function at the dielectric boundary) formalism, we were able to account for the heterogeneous influence of biological membranes in terms of a solvent-inaccessible infinite planar low-dielectric slab. The reliability of the membrane GB model was assessed by exploring the role of biological membranes in influencing the conformational changes in membrane (with melittin from bee venom), the tilt of transmembrane peptides with respect to the membrane normal (with the M2 protein from influenza A), and helix-helix interactions in membranes and the prediction of the configuration of transmembrane helical bundles (with the transmembrane domain of glycophorin A). By combining the membrane GB model with advanced computational sampling methods, like replica-exchange (REX) molecular dynamics (MD), we were able to fold and assemble helical membrane peptides correctly.

PREDICTOR@HOME: PROTEIN STRUCTURE PREDICTION VIA THE WORLD-WIDE-WEB (0026)

BTR Unit: Technological Research and Development
%BTR $: 5.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
AN, CHAHM, BS	MOLECULAR BIOLOGY
TAUFER, MICHELA, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

This core research project is directed toward the development, implementation and application of structural bioinformatics methods to protein structure prediction, loop modeling and homology modeling. During the last year we have deployed a new computational infrastructure for protein structure prediction utilizing volunteer computing resources through a world-wide-web distributed computing mechanism. This architecture is based on the BOINC (Berkeley Open Infrastructure for Network Computing) middleware (http://boinc.berkeley.edu) and the software packages CHARMM (biomolecular simulation package) and mfold (lattice-based protein folding code, part of the MMTSB Tool Set) called predictor@home (http://predictor.scripps.edu). Predictor@home is a world-community experiment and effort to use distributed world-wide-web volunteer resources to assemble a supercomputer able to predict protein structure from protein sequence. Our work is aimed at testing and evaluating new algorithms and methods of protein structure prediction in the context of the Sixth Biannual CASP (Critical Assessment of Techniques for Protein Structure Prediction) experiment. The goal is to utilize these approaches together with the immense computer power that can be harnessed through the internet and volunteers all over the world to address critical biomedical questions of protein-related diseases. Our ultimate objective in this project is to deploy these resources to provide a service component structure prediction server for the scientific community.

Subproject Progress

Progress has continued in the development and deployment of this resource.

OPTIMIZED IMPLICIT SOLVENT FORCE FIELDS (0027)

BTR Unit: Technological Research and Development
%BTR $: 2.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
CHEN, JIANHAN, BS, PHD	MOLECULAR BIOLOGY
IM, WONPIL, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

The successes and failures of various implicit solvent models in their applications to the biological problems arise in principle from their ability in maintaining delicate balance of energetics between sets of competing interactions, i.e., the solvation preference of each sidechain and the peptide backbone in aqueous bulk solution versus the strength of solvent-mediated interactions between these moieties in a complex protein environment. To what extent a GB force field can capture this delicate balance appears to be a key in the success of its applications.

Subproject Progress

Recently, we have calibrated the backbone H-bond strength as well as water-mediated sidechain-sidechain interactions by adjusting the input radii for a GB model. Since the dielectric boundary in continuum dielectric solvent models dictates all of the electrostatic and most of nonpolar solvation energetics, it is physically appropriate to optimize the input radii not only based on solvation of individual sidechains but also in consideration of solvent-mediated interactions. Our study of conformational equilibria of two helical peptides, (AAQAA)3 and ?-lac, demonstrated that the helical content can be significantly influenced with small modification of the backbone input radii. For instance, based on the explicit-solvent H-bond PMFs, a small change of the amide nitrogen input radius from 2.23 Å (the Nina's radii) to 1.95 Å, which corresponds to 0.4 kcal/mol H-bond strength difference in an alanine dipeptide model, alters the average helicity of (AAQAA)3 at 270 K from 83% to 55%, closer to the 50% value observed in experiments. In the case of ?-lac, the same modification results in about a 25% reduction in the helicity of residues 108-111, which is in accord with the experimental results that this segment of the peptide is largely unstructured in water.

TAMD FOR EFFICIENT CONFORMATIONAL SAMPLING (0029)

BTR Unit: Technological Research and Development
%BTR $: 2.000%

Investigator	Department/Non-Host Information
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
CHEN, JIANHAN, BS, PHD	MOLECULAR BIOLOGY
IM, WONPIL, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

We have recently implemented torsion angle molecular dynamics (TAMD) in CHARMM and investigated its application to augment conformational sampling of peptides and proteins. Interesting conformational changes in proteins mainly involve torsional degrees of freedom. Carrying out molecular dynamics in torsion space does not only explicitly sample the most relevant degrees of freedom, but also allows larger integration time steps with elimination of the bond and angle degrees of freedom. However, the covalent geometry needs to be fixed during internal coordinate dynamics, which can introduce severe distortions to the underlying potential surface in the extensively parameterized modern Cartesian-based protein force fields. A ``projection'' approach is extended to construct an accurate internal coordinate force field (ICFF) from a source Cartesian force field. Torsion cross-term corrections constructed from local molecular fragments, together with softened van der Waals and electrostatic interactions, are used to recover the potential surface and incorporate implicit bond and angle flexibility. MD simulations of dipeptide models demonstrated that full flexibility in both the backbone phi/psi and side chain chi angles were virtually restored. The efficacy of TAMD in enhancing conformational sampling was then further examined by folding simulations of small peptides and refinement experiments of protein NMR structures. The results showed that an increase of several fold in conformational sampling efficiency can be reliably achieved. The current study also revealed some complicated intrinsic properties of internal coordinate dynamics, beyond energy conservation, that can limit the maximum size of the integration time step and thus the achievable gain in sampling efficiency.

MODEL BUILDING & SIMULATION OF DNA & RNA AT MULTIPLE LENGTH SCALES (0006)

BTR Unit: Technological Research and Development
%BTR $: 10.000%

Investigator	Department/Non-Host Information
CASE, DAVID, AM, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

Project I of the resource is targeted at development and testing of reduced models for DNA and RNA. We have been exploring electrostatic interactions in nucleic acids, specifically by considering the application of continuum solvent models (where all the degrees of freedom of the solvent and mobile counterions are removed) to conformational problems in nucleic acids. We have shown, for the first time, that these effective solvation models provide a convincing account of structures and relative energies of DNA and RNA helices and loops. Key goals for the coming project period will be to further explore divalent metal-ion binding and to incorporate realistic low-resolution models for RNA and DNA into the CHARMM and Amber programs.

Subproject Progress

During the past year, we continued our study of the application of generalized Born (GB) and Poisson-Boltzmann (PB) models to study protein-protein association, particularly in the Ras-Raf portion of protein signaling chains. In addition to studying the convergence of numerical estimates of association energies, we analyzed the changes in flexibility upon binding in terms of molecular framework approach, called FIRST, which treats chemical bonds and H-bond interactions in a manner analgous to struts in bridges and other mechanical structures. In a computational time of about one second, FIRST identifies flexible and rigid regions in a single, static three-dimensional molecular framework, whose vertices represent protein atoms and whose edges represent covalent and non-covalent (hydrogen bond and hydrophobic) constraints and fixed bond angles within the protein. The two methods show a very good agreement with respect to the identification of changes in flexibility in both binding partners on a local scale. This implies that flexibility can be successfully predicted by identifying which bonds limit motion within a molecule and how they are coupled. In particular, as identified by MD, the _-sheet in Raf shows considerably more pronounced orientational correlations in the bound state compared to the unbound state.Similarly, FIRST assigns the beta-sheet to the largest rigid cluster of the complex. Interestingly, FIRST allows us to identify that interactions across the interface (but not conformational changes upon complex formation) result in the observed rigidification. Since regions of the beta-sheet of Raf that do not interact directly with Ras become rigidified, this also demonstrates the long-range aspect to rigidity percolation. Possible implications of the change of flexibility of the Ras-binding domain of Raf on the activation of Raf upon complex formation were assessed.

A second major effort looked at low-resolution means for docking DNA-binding proteins to their cognate DNA partners. Determining protein?DNA interactions is important for understanding gene regulation, DNA repair and chromatin structure. Unfortunately, the structures of DNA-bound complexes are often difficult to obtain experimentally, so the development of computational methods that provide good models of these complexes would be valuable. Here, we present a rigid-body docking approach using the computer program DOT. DOT performs a complete, six-dimensional search of all orientations for two rigid molecules and calculates the interaction energy as the sum of electrostatic and van der Waals terms. DOT was applied to three winged-helix transcription factors that share similar DNA-binding structural motifs but bind DNA in different ways. Docking with linear B-form DNA models accomplished several objectives; it (1) distinguished the different ways the transcription

factors bind DNA, (2) identified each protein?s DNA-binding site and the DNA orientation at the site and (3) gave at least one solution among the three best-ranked that shows the protein side chain?DNA base interactions responsible for recognition. Furthermore, the ensemble of top-ranked, docked linear B-DNA fragments indicated the DNA bending induced upon protein binding. Docking linear B-DNA to structures of the transcription factor FadR suggests that the allosteric, conformational change induced upon effector binding results in loss of the ability to bend DNA as well as loss of sequence-specific interactions with DNA. The electrostatic energy term calculated by DOT is comparable to the electrostatic binding energy calculated by Poisson?Boltzmann methods. Our results show rigid-body docking that includes a rigorous treatment of the electrostatic interaction energy can be effective in predicting protein?DNA interactions.

In a third major project, we have participated in a computationally intensive project aimed at carrying out molecular dynamics (MD) simulations including water and counterions on B-DNA oligomers containing all 136 unique tetranucleotide base sequences. This initiative was undertaken by an international collaborative effort involving nine research groups, the "Ascona B-DNA Consortium" (ABC). Calculations were carried out on the 136 cases imbedded in 39 DNA oligomers with repeating tetranucleotide sequences, capped on both ends by GC pairs and each having a total length of 15 nucleotide pairs. All MD simulations were carried out using a well defined protocol, the AMBER suite of programs, and the parm94 force field. Phase I of the ABC project involves a total of roughly 0.6 microseconds of simulation for systems containing roughly 24,000 atoms. The resulting trajectories involve 600,000 coordinate sets and represent roughly 400 GB of data. The results indicated the sequence context effects to be small for this step, but revealed that MD on DNA at this length of trajectory is subject to surprisingly persistent cooperative transitions of the sugar-phosphate backbone torsion angles ? and ?. A detailed analysis of the entire trajectory database and occurrence of various conformational substates and its impact on studies of context effects reveals a possible direct correspondence between the sequence dependent dynamical tendencies of DNA structure and the tendency to undergo transitions that ?trap? them in non-standard conformational substates. The difference in mean of the observed base pair step helicoidal parameter distribution with different flanking sequence sometimes differs by as much as 1 standard deviation, indicating that the extent of sequence effects could be significant. The observation reveals that the impact of a flexible dinucleotide such as CpG could extend beyond the immediate base pair neighbors. The results in general provide new insight into MD on DNA and the sequence dependent dynamical structural characteristics of DNA.

MULTISCALE MODELS FOR PACKAGING VIRAL DNAS (0015)

BTR Unit: Technological Research and Development
%BTR $: 4.000%

Investigator	Department/Non-Host Information
HARVEY, STEPHEN, BS, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY, GA USA
LOCKER, REBECCA C, BS, PHD	BIOLOGICAL SCIENCES GEORGIA TECHNOLOGICAL INSTITUTE, GA USA

Subproject Description

We are continuing our investigation of the packaging of double-stranded DNA into viral capsids, with particular emphasis on bacteriophage. The studies use a reduced representation model with one pseudoatom for every ten basepairs (original model), or one pseudoatom for every six basepairs (revised model). The goal is to understand the structural, kinetic, and thermodynamics of DNA packaging into viruses.

During the past year, we have completed extensive studies of three model viruses, P4, phi29 and T7. These differ in the size of their genomes (~10kb, ~20kb and ~40kb, respectively), in the size of the capsids, and in the fraction of free volume within the capsid filled by DNA. In addition, the interior of the P4 and phi29 capsids are empty, but T7 has an internal protein core that plays a role in the organization of the final packaged structure. We are preparing a paper for publication that describes the results of this work (Locker, Fuller & Harvey, in preparation).

MOLECULAR MECHANICS RESTRAINTS FOR BASE PAIRING & STACKING (0016)

BTR Unit: Technological Research and Development
%BTR $: 0.000%

Investigator	Department/Non-Host Information
HARVEY, STEPHEN, BS, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY, GA USA

Subproject Description

In building and refining nucleic acid structures, it is often desirable to enforce particular base pairing and/or base stacking interactions. Energy-based modeling programs with classical molecular mechanics force fields do not lend themselves to the easy imposition of penalty terms corresponding to such restraints, because the requirement that two bases lie in or near the same plane (pairing) or that they lie in parallel planes (stacking) cannot be easily expressed in terms of traditional interactions involving two atoms (bonds), three atoms (angles) or four atoms (torsions). We have derived expressions that define a collection of pseudo-bonds and pseudo-angles through which molecular mechanics restraints for base pairing and stacking can be imposed. We have implemented these restraints into the JUMNA package for modeling DNA and RNA structures. JUMNA scripts can specify specify base pairing with a variety of standard geometries (Watson-Crick, Hoogsteen, wobble, etc.), or with user-de fined geometries; they can also specify stacking arrangements. We have also implemented "soft core" functions to modify van der Waals and electrostatic interactions to avoid steric conflicts in particularly difficult refinements where two backbones need to pass through one another. The utility of the method was examined in a handful of test cases. The constraints could be adapted for implementation in other molecular mechanics packages.

Subproject Progress

No activity during the current year.

MULTISCALE MODELS FOR PACKAGING VIRAL RNAS (0017)

BTR Unit: Technological Research and Development
%BTR $: 5.000%

Investigator	Department/Non-Host Information
HARVEY, STEPHEN, BS, PHD	SCHOOL OF BIOLOGY GEORGIA INSTITUTE OF TECHNOLOGY, GA USA
JOHNSON, JOHN E, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

We are continuing our investigations into the three-dimensional structure of the genomic RNA of pariacoto virus (PaV). The crystal structure reveals about 35% of the RNA genome, which forms a dodecahedral cage immediately inside the capsid (Tang et al., 2001). A three-dimensional model for the RNA requires that the symmetry of the dodecahedral cage be broken to allow connection of the RNA in the outer cage with that further in the interior. The connections (which we call ¿stalactites¿) hang from the vertices of the dodecahedron. The organization of the RNA must permit a path traced by the backbone of a single RNA molecule to cover each edge of the dodecahedron twice, once in each direction to give antiparallel helices, and to drop down into the interior at each stalactite and then returns to the surface cage from the same stalactite.

An important advance of the past year has been the examination of the secondary structure of PaV RNA, in an effort to explain how the virus is able to selectively package its own RNA. The capsid proteins have nonspecific RNA-binding activity, and virus-like structures can be assembled from capsid proteins and any RNA ¿ tRNAs, rRNAs and mRNAs. If viral RNA (vRNA) is present, however, mature viruses are formed that contain only vRNA, to the exclusion of all other RNAs. The basis for this strong selectivity is not known. We hypothesize that the symmetry of the dodecahedral RNA structure is only consistent with certain classes of secondary structure, because the cage requires 30 double helices, each about 20 basepairs in length. If our hypothesis is correct, the distribution of energetically accessible vRNA secondary structures should be substantially different from those of other RNAs. Preliminary results support this hypothesis.

Subproject Progress

No work has occured on this project during the past year.

DEVELOPMENT OF EXPERIMENTAL SYSTEMS FOR COMPUTATIONAL STUDY (0008)

BTR Unit: Technological Research and Development

%BTR $: 3.000%

Investigator	Department/Non-Host Information
JOHNSON, JOHN E, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

No activity reported during the current year.

MODELING VIRUS ASSEMBLY STRUCTURE, ENERGY & THERMODYNAMICS (0002)

BTR Unit: Technological Research and Development
%BTR $: 3.000%

Investigator	Department/Non-Host Information
REDDY, VIJAY S, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
JOHNSON, JOHN E, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

This project is directed towards the specific aims: 1) develop and maintain the VIrus Particle ExploreR (VIPER), a database and website, http://mmtsb.scripps.edu/viper/, of icosahedral virus structures 2) devise the structural and computational tools to analyze virus structures.

Subproject Progress

In a significant development in the past year, we successfully ported the entire VIPER site, all the data (coordinates), related information (e.g., taxonomical) and computationally derived results, into a MySQL relational database environment. This new database, which we now call VIPERdb, is currently accessible at the URL http://viperdb.scripps.edu will replace existing VIPER site in the year 2006. Furthermore, VIPERdb contains a lot more entries (211) of capsids that include all the viral capsid entries: virus-drug complexes, mutant structures and models derived based on cryoEM reconstructions. The database was created by Dr. Craig Shepherd, as postdoctoral fellow with assistance from Mr. Ian Borelli under the supervision of Dr. Vijay Reddy.

The derived data and pictorial (graphical) descriptions for all the entries were regenerated. The derived results include, contact tables, subunit-subunit buried surface areas, solvation energies, association energies, measures of quasi-equivalence (Q-scores) and accessible surface profiles and are stored as tables in the database. Some of these properties can be plotted on graphs as a function of residue numbers using JpGraph and PHP scripts in real time in a browser window. The location of the individual contacts in the contact tables can be visualized in real time on a browser widow and interacted with using VRML plugins.

To enable the users to gain access to the VIPERdb, a relational database, a window to browse the database is provided on the website to query the database. For instance the following query submitted through such window indentifies the highly exposed cysteine residues in a particular virus (Cowpea Mosaic Virus, CpMV) and sort in the descending order based on their radial distance.

query: SELECT viper.entry_id, label_asym_id, label_comp_id, label_seq_id, eff_rad_x_sasa_min, sasa_bound, radius_aa FROM viper_residue_asa, layer, viper WHERE viper_residue_asa.entry_id =layer.entry_id AND viper_residue_asa.entry_id = viper.entry_id AND sasa_bound 0 AND viper.entry_id = '1ny7' AND (label_comp_id ='CYS')order by radius_aa desc

ENTRY_ID	LABEL_ASYM_ID	LABEL_COMP_ID	LABEL_SEQ_ID	SASA_BOUND	RADIUS_AA
1ny7	A	CYS	4	21.4	128.8
1ny7	C	CYS	119	8.4	117.8
1ny7	B	CYS	295	20.2	112.5
1ny7	C	CYS	132	2.7	112.3
1ny7	B	CYS	355	34.7	108.5

1ny7 C CYS 177 1.5 106.7
1ny7 B CYS 187 85.9 105.9

The power and utility of such a database is immense for comparing residue properties across different viruses within a family and even across families. Such comparisons are currently on going

We are extremely happy to report that VIPERdb was featured in the netwatch column in 18 Feb. 2005 issue of science magazine.

Another front of significant development in the past year is EMDB, a website (http://mmtsb1.scripps.edu/viper/em_index.php) of cryo-EM structures (densities) of icosahedral viruses. The EMDB is being developed by Dr. Padmaja Natarajan under the supervision of Professor Jack Johnson as extension and augment the VIPER database. There are currently over 40 EM reconstructions for which both electron densities and structure factor amplitudes and phases are available through this web site. A review article on how to use VIPER site accepted for publication in Nature reviews in Microbiology.

ANALYSIS OF PROTEIN-PROTEIN INTERACTIONS AND QUASI-EQUIVALENCE IN VIRUS CAPSIDS (0003)

BTR Unit: Technological Research and Development

%BTR $: 4.000%

Investigator	Department/Non-Host Information
REDDY, VIJAY S, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
SHEPHERD, CRAIG M, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

Viral capsids are composed of multiples copies of one or a few coat protein subunits. Knowledge on the extent of interactions of a capsid protein subunit in native and sometimes in non-native capsid environments will be of importance in understanding the relationship between the protein building block and the capsid size and architecture. Furthermore, the extent of similarity in the bonding and interactions of the subunits would point to the quasi-equivalences of subunits in viral capsids. The knowledge on the relationship between the type of the protein and the resulting capsid would aid in designing new and novel capsids.

The objective of this project is to come up with a measure/property that would quantify the extent to which a subunit is involved in the protein-protein interactions. To this end, we came up with fraction buried surface area of the subunit, which we call protein-protein interaction (PPI) index, as a measure of extent of protein-protein interactions correspond to a subunit in the capsid environment.

Subproject Progress

The solvent accessible surface areas (SASA) were calculated using the CHARMM molecular mechanics software package, using the CHARMM19 set of united atom parameters and a solvent probe radius of 1.4 Å. Buried surface area (BSA) was calculated by subtracting the SASA values of a subunit in the context of the capsid from the SASA of the same subunit in isolation. The SASA of the reference subunit in the context of the capsid was determined using a CHARMM stream file (script), which loops over and subtracts the BSAs due to every nearest subunit from the SASA of the isolated reference subunit. For capsids with T1 the average values of protein-protein interaction (PPI) indices of the distinct subunits in each capsid are plotted with the standard deviations are shown as error bars.

In the case of T=1 capsids, the PPI indicies show an increase proportional to the molecular weight (size) of the coat protein subunit, where as the indicies for the subunits of the quasi-equivalent capsids interestingly are rather independent of the size. Remarkable outcome of such a study was the distinct clustering of ?T=2? capsids with low PPI indicies and larger subunit molecular weights, which uniquely distinguish them as the thin shelled capsids with lager surface area per unit molecular weight. Similar analyses done as a function of capsid diameter also result in analogous results.

The variation (residual) in PPI indices of subunits of quasi-equivalent capsids was used as a measure to estimate extent of quasi-equivalence (1-residual) in capsids with T number 1. This study has been published in the proteins in the year 2005.

DEVELOPING TOOLS FOR ANALYSIS OF VIRUS STRUCTURES (0004)

BTR Unit: Technological Research and Development
%BTR $: 4.000%

Investigator	Department/Non-Host Information
REDDY, VIJAY S, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

The VIPER tool set contains a number analysis tools that have been developed over the past few years to analyze virus structures particularly in an automatic fashion, which also can be employed for the analysis of multimeric proteins. The objective of this project is to organize the VIPER toolset for distribution on the same lines as the MMTSB Tool Set.

These structural tools (Table 1) have been modified to generate the derived results/data to be uploaded to into a VIPERdb. In a significant addition to the tools we developed a new an algorithm and tool that automatically generates the PDB_to_VIPER matrix, which transforms the .pdb coordinates into VIPER convention. This tool is a hugh step forward in removing the bottleneck in performing the VIPER analysis in an automatic fashion.

Subproject Progress

All the resultant analyses from these tools can be done currently through the web interfaces from the VIPERdb site. Efforts are on going to generate a standalone tools set for the distribution to the structural virology community.

In addition a number of web-based utilities such as oligomer generator, gallery maker, contact finder, etc., which now operate by interacting with the relational database are also available at the VIPERdb (http://viperdb.scripps.edu) site.

Table 1. Various analysis tools available currently as part of the VIPER toolset.

S.No.	Name of the utility*	Function of the utility
1	process_pdb.pl	reads in pdb files and outputs CHARMM coordinate files
2	gen_hydr.stream	generates hydrogens and outputs .chr files
3	gen_segments.stream	Generates duplicate chains of subunits in the icos. asymmetric unit.
4	list_contacts.stream	Given NCS matrices & chains generates a list of possible subunit pairs
5	select_contacts.stream	Identifies a subset of the subunits pairs that are truly in contact
6	rotate_generic.stream	Transforms the coordinates by a chosen NCS matrix
7	coor_dist.stream	Identifies the number of residues pairs in contact, cutoff limit 8.5A
8	count.pl	Counts the number of residue pairs in contact that have been identified
9	select_unique.pl	Selects the unique subunit pairs that are in contact.
10*	gen_contact_table.stream	identifies and organizes residue pairs in contact at the subunit interfaces into columns
11*	qscore.stream	Calculate similarity score based on common residue pairs at the subunit interfaces
12*	bsa_solv_energy.stream	Calculates the solvation energies of subunit pairs based on buried ASAs.
13*	ppi.stream	Calculates protein-protein interaction index
14*	select_insert_site.stream	Identifies the potential sites for the insertion of loops on the viral surface.
15*	net_charge.stream	Caclculates net charge on the virus surface.
16*	pdbtoviper.c	Generates PDB to VIPER matrix

Files with ?.pl? extension are perl scripts and ?.stream? are CHARMM stream (script) files. Utilities identified by the ?*? refers to the tools that calculates various derived properties/results.

SCORING OF MODELS FITTED INTO CRYO EM DENSITIES (0005)

BTR Unit: Technological Research and Development
%BTR $: 0.000%

Investigator	Department/Non-Host Information
REDDY, VIJAY S, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
JOHNSON, JOHN E, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

Problem and significance: Validate and score models against the experimental cryo-EM data. Towards this goal, we used real and reciprocal space methods to score models against the cryo-EM data. Reciprocal space methods: Amplitudes and phases from the experimental EM maps were used as restraints to refine models into cryo-EM density. The crystallographic refinement program X-PLOR [1] was used to do the refinement. Trials using the NWV procapsid data revealed that reciprocal space based methods have greater radius of convergence than the real space methods. Real space methods: The program RSREF [2] was used refine models into cryo-EM density. This program is beneficial in the cases, where only a partial density is used to refine the corresponding model. But the radius of convergence is turned out to be rather poor in the real space methods. The model needed to be positioned visually before the start of refinement. Reciprocal space refinement was successfully used to follow structural transformations starting from the model for one of the states reflecting the higher radius of convergence. While the same could not be achieved through RSREF, a real space method.We are currently in the process of comparing models derived using the other programs such as SITUS [3].

Subproject Progress

No activity in this project area this year.

MULTIRESOLUTION SAMPLING METHODS FOR PROTEIN & PEPTIDE CONFORMATIONAL SPACE (0019)

BTR Unit: Technological Research and Development

%BTR $: 19.000%

Investigator	Department/Non-Host Information
SKOLNICK, JEFFREY, BA, PHD	CENTER FOR EXCELLENCE IN BIOIN SUNY, BUFFALO, NY USA

Subproject Description

This core project focuses on the development, validation, and application of hybrid models designed to explore various aspects of protein and peptide folding, ligand docking and thermodynamics. In that regards, the recently developed TASSER algorithm was applied to predict the tertiary structures of all CASP6 targets. TASSER is a hierarchical approach that consists of template identification by the threading program PROSPECTOR_3, followed by tertiary structure assembly via rearranging continuous template fragments. For all 90 CASP targets/domains, PROSPECTOR_3 generates initial alignments with an average root-mean-square deviation to native (RMSD) of 8.4 Å with 79% coverage. After TASSER reassembly, the average RMSD decreases to 5.4 Å over the same aligned residues; the overall cumulative TM-score increases from 39.44 to 52.53. Despite significant improvements over the initial template alignment observed in all target categories, the overall quality of the final models is dictated by the threading templates quality. The average TM-score of TASSER models in the three categories are 0.79 (CM, 43 targets/domains), 0.47 (FR, 37 targets/domains), and 0.30 (NF, 10 targets/domains). This highlights the need to develop novel (or improved) approaches to identify very distant targets as well as better NF algorithms.

The ability of the combined AMBER/GBSA potential to identify native from misfolded structures as well as the correlation with native-likeness was assessed on a 30 protein test set, a subset of the PDB200 benchmark set of nonhomologous proteins representative of the PDB between 41 and 200 residues. For each protein, 14,000 decoys, previously generated by TASSER, were used. When the decoys were only locally minimized with AMBER/GBSA, nearly all of the native structures from the set are recognized as the lowest energy structures, with an average native-decoy energy gap of around 10% of the total native energy. The few failures involved proteins that were crystallized as a part of a complex. In a more difficult test, the decoys were subject to 100 ps Molecular Dynamics relaxation at 300° K followed by minimization. Now, the native structure is not the lowest energy one, and the AMBER/GBSA energy does not generally exhibit a correlation with native-likeness as assessed by the root-mean-square-deviation from native. In the coming year, the entire PDB200 set will be evaluated and an optimization of the AMBER/GBSA force field will be performed.

VIRUS SWELLING STUDIED VIA ELASTIC NETWORK NORMAL MODES (0001)

BTR Unit: Technological Research and Development
%BTR $: 3.000%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY

Subproject Description

During the past five years there has been a tremendous growth of the number of virus structures solved by X-ray crystallography and cryo electron microscopy (EM) experiments, including virus capsids adopting several functionally relevant conformations. In addition, with time resolved small angle scattering experiments dynamics of viruses can now be monitored and for some systems several intermediate have been observed. These large-scale rearrangements are an integral part of the life cycle of the virus and have been observed in the maturation process, during assembly and may be involved in the release of RNA during infection. Disrupting the assembly or the maturation processes in viruses may be potential targets for antiviral activities. Thus, the description, prediction and exploration of such conformational changes for viruses are essential to understand their mechanism of activity and should be helpful in the development of new drugs.

Subproject Progress

Therefore, the mechanical properties and putative dynamical fluctuations of a variety of viral capsids comprising different sizes and quasi-equivalent symmetries were analyzed by performing normal mode analysis using the elastic network model. The expansion of the capsid to a swollen state is studied using normal modes and is compared with the experimentally observed conformational change for three of the viruses for which experimental data exists. A combination of one or two normal modes was shown to capture remarkably well the overall translation that dominates the motion between the two conformational states, and reproduces the overall conformational change. For all of the viral capsids it is observed that the nature of the modes is different. In particular for the T=7 virus, HK97, for which the shape of the capsid changes from spherical to faceted polyhedra, two modes are necessary to accomplish the conformational transition. In addition, extension of this study to viral capsids with other T numbers reveals the similarities and differences in the features of virus capsid conformational dynamics. In particular, the pentamers generally have higher flexibility and propensity to move freely from the other capsomers, which facilitates the shape adaptation that may be important in the viral life cycle.

Recently we have also started to explore the effect of RNA on the swelling process of viruses. We are currently studying the Satellite Tobacco Mosaic Virus (STMV) which is a T=1 icosahedral virus. About 60 % of its genomic material has been identified by X-ray crystallography. It has been observed experimentally that upon change in pH or RNA degradation, STMV adopts a swollen form (its radius increases by 5 Å). Using the lowest frequency - non-degenerate - normal mode to model the swelling process, strain energy analysis reveals that the swelling is stiffer when RNA is present which could potentially explain why STMV adopts a conformation with a larger radius when RNA is degraded.

NMFF - A NEW PROGRAM FOR FLEXIBLE FITTING OF ATOMIC STRUCTURES INTO EM MAPS (0020)

BTR Unit: Technological Research and Development
%BTR $: 5.000%

Investigator	Department/Non-Host Information
TAMA, FLORENCE, BS, PHD	MOLECULAR BIOLOGY
BROOKS, CHARLES, BS, PHD	MOLECULAR BIOLOGY
MIYASHITA, OSAMU, BS, PHD	PHYSICS UCSD, CA USA

Subproject Description

Cryo-EM is emerging as a primary tool not only for the elucidation of structural information but also for the examination of dynamical properties of supramolecular assemblies. To complement emerging information from EM, it is tempting to use existing X-ray data to construct atomic models. As large assemblies often undergo large functional rearrangements, the conformations of the available X-ray structures may not correspond to the conformation apparent in the cryo-EM data. In such cases, the fitting can be rather complicated, since not only the correct orientation needs to be found but conformational rearrangements must also be considered. However, current approaches deform the molecule in an ad-hoc way. Therefore, we have introduced a novel method for the quantitative flexible docking of a high-resolution structure into low-resolution maps of macromolecular complexes from EM that takes into account the conformational flexibility of biological systems.

Subproject Progress

The algorithm has been tested on several tests cases and we have demonstrated that the results are robust. We have also been working in close collaboration with several groups to fit high-resolution structures into experimental maps of variety of systems. The discussions with our collaborators, who will be the potential users of the algorithm, have been helpful to prepare a released version of NMFF. The release of NMFF for general use within the scientific community is planned for the summer 2005.

PUBLISHED: ABSTRACTS, BOOKS, & JOURNALS

‡ Center Cited

SPIDs	Reference
Books	
0004, 0002, 0005, 0003	‡ Rao, ALN & Reddy, VS (2004).Architecture of plant viruses. in Handbook of Plant Virology, eds. Khan, JD & Dijkstra, J (Haworth Press)
0004, 0002, 0003	‡ Reddy, VS, Natarajan, P, Lander, G, Qu, C, Brooks, CL, III & Johnson, JE (2004).Virus Particle Explorer (VIPER): A repository of virus capsid structures in Conformational proteomics of macromolecular Architecture, eds. Cheng, RH & Hammer, L (World Scientific Pub Co., Singapore), Vol. Chpt. 20, pp. 403-12
0001	‡ Tama, F & Brooks, CL, III (2004).Unveiling molecular mechanisms of biological functions in large macromolecular assemblies using elastic network normal mode analysis in Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems, eds. Cui, Q & Bahar, I (CRC Press)
Journals	
0006	‡ Beveridge David L; Barreiro Gabriela; Byun K Suzie; Case David A; Cheatham Thomas E; 3rd Dixit Surjit B; Giudice Emmanuel; Lankas Filip; Lavery Richard; Maddocks John H; Osman Roman; Seibert Eleanore; Sklenar Heinz; Stoll Gautier; Thayer Kelly M; Varnai Péter; Young Matthew A. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. Biophys J 87 6 2004 Dec pp.3799-813.
0010	‡ Bindewald Eckart; Skolnick Jeffrey. A scoring function for docking ligands to low-resolution protein structures. J Comput Chem 26 4 2005 Mar pp.374-83.
0028	‡ Chen Jianhan; Im Wonpil; Brooks Charles L; 3rd . Refinement of NMR structures using implicit solvent and advanced sampling techniques. J Am Chem Soc 126 49 2004 Dec 15 pp.16038-47.
0028	‡ Chen Jianhan; Won Hyung-Sik; Im Wonpil; Dyson H Jane; Brooks Charles L; 3rd . Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. J Biomol NMR 31 1 2005 Jan pp.59-64.
0006	‡ Dixit, SB, Beveridge, DL, Case, DA, T.E. Cheatham, I, Giudice, E, Lankas, F, Lavery, R, Maddocks, JH, Osman, R, Sklenar, H, Thayer, KM & Varmai, P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II. Sequence contect effects on the dynamical structures of the 10 unique dinucleotide steps. Biophys J (2005).
0021	‡ Falke Scott; Tama Florence; Brooks Charles L; 3rd Gogol Edward P; Fisher Mark T. The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. J Mol Biol 348 1 2005 Apr 22 pp.219-30.
0012	‡ Feig Michael; Brooks Charles L; 3rd . Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol 14 2 2004 Apr pp.217-24.
0012	‡ Feig Michael; Im Wonpil; Brooks Charles L; 3rd . Implicit solvation based on generalized Born theory in different dielectric environments. J Chem Phys 120 2 2004 Jan 8 pp.903-11.
0011, 0019	‡ Feig Michael; Karanicolas John; Brooks Charles L; 3rd . MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 22 5 2004 May pp.377-95.
0011	‡ Feig Michael; Onufriev Alexey; Lee Michael S; Im Wonpil; Case David A; Brooks Charles L; 3rd . Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25 2 2004 Jan 30 pp.265-84.

SPIDs	Reference
0012	‡ Ferrara Philippe; Gohlke Holger; Price Daniel J; Klebe Gerhard; Brooks Charles L; 3rd . Assessing scoring functions for protein-ligand interactions. J Med Chem 47 12 2004 Jun 3 pp.3032-47.
0012	‡ Gohlke Holger; Case David A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 25 2 2004 Jan 30 pp.238-50.
0019	‡ Gohlke Holger; Kuhn Leslie A; Case David A. Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. Proteins 56 2 2004 Aug 1 pp.322-37.
0012	‡ Im Wonpil; Brooks Charles L; 3rd . De novo folding of membrane proteins: an exploration of the structure and NMR properties of the fd coat protein. J Mol Biol 337 3 2004 Mar 26 pp.513-9.
0012, 0019	‡ Im Wonpil; Brooks Charles L; 3rd . Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. Proc Natl Acad Sci U S A 102 19 2005 May 10 pp.6771-6.
0009	‡ Karanicolas, J & Brooks, CL, III. An evolution of minimalist models for protein folding: from the behavior of protein-like polymers to protein function. Biosilico 2: 127-33 (2004)
0010	‡ Kolinski, A & Skolnick, J. Reduced Models of Proteins and Their Applications. Polymer 45: 511-24 (2004).
0026	‡ Kondo, D, Taufer, M, Brooks, CL, III, Casanova, H & Chien, A (2004) in Proceedings of IPDPS 2004, IEEE/ACM International Parallel and Distributed Processing Symposium, Santa Fe, New Mexico).
0006	‡ LaMarque Jaclyn C; Le Thuc-Vy L; Harvey Stephen C. Packaging double-helical DNA into viral capsids. Biopolymers 73 3 2004 Feb 15 pp.348-55.
0027	‡ Mackerell Alexander D; Jr Feig Michael; Brooks Charles L; 3rd . Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25 11 2004 Aug pp.1400-15.
0027	‡ MacKerell Alexander D; Jr Feig Michael; Brooks Charles L; 3rd . Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 126 3 2004 Jan 28 pp.698-9.
0026	‡ Natrajan, A, Crowley, M, Wilkins-Diehr, N, Humphrey, MA, Fox, AD, Grimshaw, AS & Brooks, CL, III. Studying protein folding on the Grid: experiences using CHARMM on NPACI resources under Legion. Concurrency and Computation: Practice and Experience 16: 385-97 (2004).
0006	‡ Roberts Victoria A; Case David A; Tsui Vickie. Predicting interactions of winged-helix transcription factors with DNA. Proteins 57 1 2004 Oct 1 pp.172-87.
0003	‡ Shepherd Craig M; Reddy Vijay S. Extent of protein-protein interactions and quasi-equivalence in viral capsids. Proteins 58 2 2005 Feb 1 pp.472-7.
0001	‡ Tama Florence; Brooks Charles L; 3rd . Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. J Mol Biol 345 2 2005 Jan 14 pp.299-314.
0014	‡ Tama Florence; Feig Michael; Liu Jun; Brooks Charles L; 3rd Taylor Kenneth A. The requirement for mechanical coupling between head and S2 domains in smooth muscle myosin ATPase regulation and its implications for dimeric motor function. J Mol Biol 345 4 2005 Jan 28 pp.837-54.
0020	‡ Tama Florence; Miyashita Osamu; Brooks Charles L; 3rd . Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. J Mol Biol 337 4 2004 Apr 2 pp.985-99.

SPIDs	Reference
0020	‡ Tama Florence; Miyashita Osamu; Brooks Charles L; 3rd . Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J Struct Biol 147 3 2004 Sep pp.315-26.
0006	‡ Tihova Mariana; Dryden Kelly A; Le Thuc-vy L; Harvey Stephen C; Johnson John E; Yeager Mark; Schneemann Anette. Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. J Virol 78 6 2004 Mar pp.2897-905.
0023	‡ Trylska Joanna; Konecny Robert; Tama Florence; Brooks Charles L; 3rd McCammon J Andrew. Ribosome motions modulate electrostatic properties. Biopolymers 74 6 2004 Aug 15 pp.423-31.
0020	‡ Wriggers, W, Chacon, P, Kovacs, JA, Tama, F & Birmanns, S. Topology representing neural networks reconcile biomolecular shape, structure, and dynamics. Neurocomputing 56: 365-79 (2004)

Published This Year		Cited	Not Cited	Total
Books		3	0	3
Journals		31	0	31
	Total	34	0	34

IN PRESS: ABSTRACTS, BOOKS, & JOURNALS

‡ BTR Cited

SPIDs	Reference
Journal	
0011, 0012, 0	‡ Case, DA, Cheatham, T, Darden, T, Gohlke, H, Luo, R, Merz, KM, Jr., Onufriev, A, Simmerling, C, Wang, B & Woods, R. The Amber Biomolecular Simulation Programs. J Comp Chem in the press (2005)
0028, 0029	‡ Chen, J, Im, W & Brooks C. L., III. Torsion Angle Molecular Dynamics for Efficient Sampling of Protein Conformations. J Comp Chem in the press (2005)
0027	‡ Im, W, Chen, J & Brooks, CL, III (2005) in Peptide H-Bonds and Peptide Solvation, eds. Baldwin, RE & Baker, DJ (Elsevier), Vol. in the press
0006	‡ Mears, JA, Stagg, SM & Harvey, SC (2005) in Handbook of RNA Biochemistry, eds. Westhof, E, Binereif, S, Schön, A & Hartmann, R (Wiley-VCH, Weinheim), Vol. in the press
0004, 0002, 0	‡ Natarajan, P, Lander, G, Shepherd, CM, Reddy, VS, Brooks, CL, III & Johnson, JE. Exploring Icosahedral Virus Structures with VIPER. Nature Rev Micro Biol in the press (2005)
0004, 0002, 0	‡ Reddy, VS & Johnson, JE. Structure derived insights into virus assembly. Adv Virus Res in the press (2004)
0010	‡ Skolnick, J & Zhang, Y (2005) in Systems Biology, ed. Rigoutsos, I (Springer-Verlag, Vol. in the press
0026	‡ Taufer, M, An, C, Kerstens, A & Brooks, CL, III (2005) in Fourth IEEE International Workshop on High Performance Computational Biology (HiCOMB 2005), Denver, Colorado, Vol. in the press
0026	‡ Taufer, M, Anderson, D, Cicotti, P & Brooks, CL, III (2005) in 14th Heterogeneous Computing Workshop (HCW 2005), Denver, Colorado), Vol. in the press
0026	‡ Taufer, M, Crowley, M, Price, D, Chien, A & Brooks, CL, III. Study of an Accurate and Fast Protein-Ligand Docking Algorithm based on Molecular Dynamics. Concurrency and Computation: Practice and Experience in the press (2004)
0023	‡ Trylska, J, McCammon, JA & Brooks, CL, III. Exploring Assembly Energetics of the 30S Ribosomal Subunit using an Implicit Solvent Approach. J Am Chem Soc in the press (2005)

In Press This Year		**Cited**	**Not Cited**	**Total**
Journals		11	0	11
	Total	11	0	11

SOURCE OF INVESTIGATORS' SUPPORT

FEDERAL

Investigator Organization	Grant/Contract	Total Funding	SPIDs
FEDERAL - NON PHS			
BROOKS, CHARLES			
NSF	MCB0413858	$155,290	0010, 0027
NSF	PHY0216576	$89,044	0009, 0012, 0020, 0023, 0024, 0026
NSF	PHY0225630	$170,676	0009, 0012, 0020, 0023, 0024, 0026
REDDY, VIJAY S			
DOD	X81XWH042007	$82,834	0002, 0003, 0004
SKOLNICK, JEFFREY			
DOD	1U54AI057158	$80,000	0019
	FEDERAL - NON PHS	**$577,844**	

FEDERAL

Investigator Organization	Grant/Contract	Total Funding	SPIDs
FEDERAL - PHS			
BROOKS, CHARLES			
NIH	2R01GM037554-18 A2	$321,880	
NIH	5R01GM048807-13	$334,182	
CASE, DAVID			
NIH	2R01GM039914-16 A2	$232,375	
NIH	5R01GM057513-07	$346,444	
NIH	5R01GM045811-15	$281,550	
NIH	3R01GM039914-15S 1	$66,829	
FRANK, JOACHIM			
NIH	5R37GM029169-24	$433,338	
NIH	5P41RR001219-23	$819,247	
NIH	5R01GM055440-08	$184,159	
GAO, JIALI			
NIH	5R01GM046736-13	$247,344	
HARVEY, STEPHEN			
NIH	5R01GM070785-02	$245,016	
NIH	5R01GM053827-10	$186,500	
JOHNSON, JOHN E			
NIH	2R01AI040101-10	$376,797	
NIH	5R37GM034220-22	$421,330	
NIH	5R01EB000432-03	$950,962	
NIH	5R01GM054076-10	$422,325	
MACKERELL, ALEXANDER			
NIH	3R01CA095200-03S 1	$33,363	
NIH	5R01GM051501-08	$218,017	
NIH	5R01CA095200-03	$231,289	
MCCAMMON, JAMES			
NIH	5R01GM031749-23	$305,835	
SKOLNICK, JEFFREY			
NIH	2R01GM048835-12	$247,275	
NIH	5R01GM037408-16	$253,057	
SMITH, THOMAS			
NIH	5R01GM010704-42	$280,728	
TAYLOR, KENNETH			
NIH	1S10RR020919-01	$240,300	
NIH	5R01AR047421-05	$208,050	
NIH	5R01GM030598-20	$249,410	

FEDERAL

Investigator Organization	Grant/Contract	Total Funding	SPIDs
	FEDERAL - PHS	**$8,137,602**	

RESOURCE SUMMARY: SUBPROJECTS

The following only includes information associated with subprojects.

	Collab	Dissem & Training	Services	Tech R & D	Total (Excludes duplicates)
Number of Subprojects	7	2	2	17	28
Number of Investigators	15	6	3	15	29
Number of Published	5	2	1	27	34
Number In Press	2	1	0	10	11
%Non-AIDS Dollars	7.000%	12.000%	3.000%	78.000%	100.000%
Total Percent of Funds Awarded	7.000%	12.000%	3.000%	78.000%	100.000%
Total IDEA dollars	0	0	0	0	100

RESOURCE SUMMARY: ADMINISTRATIVE

	On Subprojects	Not On Subprojects
Others	29	13

FEDERAL		
NON-PHS		
DOD		$162,834
NSF		$415,010
	NON-PHS	**577,844**
PHS		
AI		$376,797
AR		$208,050
CA		$264,652
EB		$950,962
GM		$5,277,594
RR		$1,059,547
	PHS	**8,137,602**
	TOTAL SUPPORT	**$8,715,446**

GEOGRAPHICAL USAGE BY INVESTIGATORS AT NON-HOST INSTITUTIONS

Foreign Investigators by Country		**1**
NEW ZEALAND	1	
USA Investigators by State		**24**
CA	7	
FL	2	
GA	3	
KS	3	
MD	1	
MI	1	
MN	1	
MO	1	
NJ	1	
NY	3	
TX	1	
		25

RESEARCH PROGRESS

Executive Summary

The primary goal of efforts in The Center for **M**ultiscale **M**odeling **T**ools in **S**tructural **B**iology (MMTSB) is the development of tools and methods to aid the structural biology community in exploring structure/function relationships in biomedically important molecules. To this end we have focused efforts in the development of software to permit: (i) the exploration of very large scale motions accompanying protein folding to be sampled using hybrid low resolution lattice-based Monte Carlo and atomic resolution molecular dynamics/mechanics based force field methods; (ii) molecular mechanical models of large nucleic and ribonucleic acid structures to be developed at low and high resolution from minimal experimental information and chemical structural data; (iii) rearrangements of macromolecular assemblies using methods of elastic network normal modes; (iv) electrostatic and association energies for large virus assemblies to be computed and categorized (v) techniques to automate the development of atomic-level structure for low resolution experimental data, including that from cryo-electron microscopy and minimal NMR restraints. Our progress during the past year has been significant, with web-based resources for protein folding and loop prediction coming on-line, revisions and updating of our web-based virus web site, making these community-used resources broader and more user friendly, and new developments in fitting of atomic-level structural models to low-resolution experimental data.

We have also made progress in addressing questions of large-scale motions in virus particles, associated with virus maturation. Our development of dynamical models for entire viral capsids, the ribosome, myosin and important enzyme systems, using simplified elastic network normal mode analysis, has been shown to capture the key features of very large-scale conformational changes associated with physical and functional processes in these systems. These methods too have been extended to enable the flexible fitting of atomic structures into cryo-EM density maps. We have developed quantitative measures of quasi-symmetry, called the protein-protein interaction index (PPI), which are now being used to explore relationships between virus capsid structures from different organisms in an effort to identify common features operative in families of related virus structures. In the area of protein folding and structure prediction, great strides have been made in the development of algorithms and models for protein fold prediction and scoring, including the development and implementation of a www-distributed platform for our Center's participation in the CASP6 exercise in blind protein structure prediction. Advances in sampling methods, including implementation of torsion angle molecular dynamics in CHARMM (TAMD), new tools and new solvation models have contributed to this exciting evolving area of structural informatics.

Finally, our training and dissemination/service efforts during the ninth year have been focused in three primary areas. The first is the refinement of our web-based server for virus assembly and structure analysis and the maturation of our MMTSB Tool Set for protein and nucleic acid modeling and structure prediction. The second has been directed toward enhancements in methods for flexible docking of molecular structures into cryo-EM data; this has been bolstered tremendously by critical collaborations with field-leading EM microscopists. Lastly, we have developed and taught workshops in building/fitting structural models into cryo-EM data and in the application of structural modeling and prediction approaches to protein structure prediction.

Progress on Individual Projects

Modeling Nucleic Acids and their Complexes at Multiple Scales of Resolution

Solvation, molecular dynamics, RNA/DNA modeling (Case):

Project I of the resource is targeted at development and testing of reduced models for DNA and RNA. We have been exploring electrostatic interactions in nucleic acids, specifically by considering the application of continuum solvent models (where all the degrees of freedom of the solvent and mobile counterions are removed) to conformational problems in nucleic acids. We have shown, for the first time, that these effective solvation models provide a convincing account of structures and relative energies of DNA and RNA helices and loops. Key goals for the coming project period will be to further explore divalent metal-ion binding and to incorporate realistic low-resolution models for RNA and DNA into the CHARMM and Amber programs.

During the past year, we continued our study of the application of generalized Born (GB) and Poisson-Boltzmann (PB) models to study protein-protein association, particularly in the Ras-Raf portion of protein signaling chains. In addition to studying the convergence of numerical estimates of association energies, we analyzed the changes in flexibility upon binding in terms of molecular framework approach, called FIRST, which treats chemical bonds and H-bond interactions in a manner analgous to struts in bridges and other mechanical structures. In a computational time of about one second, FIRST identifies flexible and rigid regions in a single, static three-dimensional molecular framework, whose vertices represent protein atoms and whose edges represent covalent and non-covalent (hydrogen bond and hydrophobic) constraints and fixed bond angles within the protein. The two methods show a very good agreement with respect to the identification of changes in flexibility in both binding partners on a local scale. This implies that flexibility can be successfully predicted by identifying which bonds limit motion within a molecule and how they are coupled. In particular, as identified by MD, the _-sheet in Raf shows considerably more pronounced orientational correlations in the bound state compared to the unbound state.
Similarly, FIRST assigns the beta-sheet to the largest rigid cluster of the complex. Interestingly, FIRST allows us to identify that interactions across the interface (but not conformational changes upon complex formation) result in the observed rigidification. Since regions of the beta-sheet of Raf that do not interact directly with Ras become rigidified, this also demonstrates the long-range aspect to rigidity percolation. Possible implications of the change of flexibility of the Ras-binding domain of Raf on the activation of Raf upon complex formation were assessed.

A second major effort looked at low-resolution means for docking DNA-binding proteins to their cognate DNA partners. Determining protein–DNA interactions is important for understanding gene regulation, DNA repair and chromatin structure. Unfortunately, the structures of DNA-bound complexes are often difficult to obtain experimentally, so the development of computational methods that provide good models of these complexes would be valuable. Here, we present a rigid-body docking approach using the computer program DOT. DOT performs a complete, six-dimensional search of all orientations for two rigid molecules and calculates the interaction energy as the sum of electrostatic and van der Waals terms. DOT was applied to three winged-helix transcription factors that share similar DNA-binding structural motifs but bind DNA in different ways. Docking with linear B-form DNA models accomplished several objectives; it (1) distinguished the different ways the transcription factors bind DNA, (2) identified each protein's DNA-binding site and the DNA orientation at the site and (3) gave at least one solution among the three best-ranked that shows the protein side chain–DNA base interactions responsible for recognition. Furthermore, the ensemble of top-ranked, docked linear B-DNA fragments indicated the DNA bending induced upon protein binding. Docking linear B-DNA to structures of the transcription factor FadR suggests that the allosteric, conformational change induced upon effector

binding results in loss of the ability to bend DNA as well as loss of sequence-specific interactions with DNA. The electrostatic energy term calculated by DOT is comparable to the electrostatic binding energy calculated by Poisson–Boltzmann methods. Our results show rigid-body docking that includes a rigorous treatment of the electrostatic interaction energy can be effective in predicting protein–DNA interactions.

In a third major project, we have participated in a computationally intensive project aimed at carrying out molecular dynamics (MD) simulations including water and counterions on B-DNA oligomers containing all 136 unique tetranucleotide base sequences. This initiative was undertaken by an international collaborative effort involving nine research groups, the "Ascona B-DNA Consortium" (ABC). Calculations were carried out on the 136 cases imbedded in 39 DNA oligomers with repeating tetranucleotide sequences, capped on both ends by GC pairs and each having a total length of 15 nucleotide pairs. All MD simulations were carried out using a well defined protocol, the AMBER suite of programs, and the parm94 force field. Phase I of the ABC project involves a total of roughly 0.6 microseconds of simulation for systems containing roughly 24,000 atoms. The resulting trajectories involve 600,000 coordinate sets and represent roughly 400 GB of data. The results indicated the sequence context effects to be small for this step, but revealed that MD on DNA at this length of trajectory is subject to surprisingly persistent cooperative transitions of the sugar-phosphate backbone torsion angles α and γ. A detailed analysis of the entire trajectory database and occurrence of various conformational substates and its impact on studies of context effects reveals a possible direct correspondence between the sequence dependent dynamical tendencies of DNA structure and the tendency to undergo transitions that "trap" them in non-standard conformational substates. The difference in mean of the observed base pair step helicoidal parameter distribution with different flanking sequence sometimes differs by as much as 1 standard deviation, indicating that the extent of sequence effects could be significant. The observation reveals that the impact of a flexible dinucleotide such as CpG could extend beyond the immediate base pair neighbors. The results in general provide new insight into MD on DNA and the sequence dependent dynamical structural characteristics of DNA.

Publications supported by the resource

Beveridge, DL, Barreiro, G, Byun, KS, Case, DA, Cheatham, TE, 3rd, Dixit, SB, Giudice, E, Lankas, F, Lavery, R, Maddocks, JH, Osman, R, Seibert, E, Sklenar, H, Stoll, G, Thayer, KM, Varnai, P & Young, MA. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* **87:** 3799-813 (2004).

Feig, M, Onufriev, A, Lee, MS, Im, W, Case, DA & Brooks, CL, III. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comp Chem* **25:** 265-84 (2004).

Gohlke, H, Kuhn, LA & Case, DA. Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* **56:** 322-37 (2004).

Gohlke, H & Case, DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comp Chem* **25:** 238-50 (2004).

Roberts, VA, Case, DA & Tsui, V. Predicting interactions of winged-helix transcription factors with DNA. *Proteins* **57:** 172-87 (2004).

Case, DA, Cheatham, T, Darden, T, Gohlke, H, Luo, R, Merz, KM, Jr., Onufriev, A, Simmerling, C, Wang, B & Woods, R. The Amber Biomolecular Simulation Programs. *J Comp Chem* **in the press** (2005).

Multiscale models for packaging viral DNAs (Harvey):

We are continuing our investigation of the packaging of double-stranded DNA into viral capsids, with particular emphasis on bacteriophage. The studies use a reduced representation model with one

pseudoatom for every ten basepairs (original model), or one pseudoatom for every six basepairs (revised model). The goal is to understand the structural, kinetic, and thermodynamics of DNA packaging into viruses.

During the past year, we have completed extensive studies of three model viruses, P4, phi29 and T7. These differ in the size of their genomes (~10kb, ~20kb and ~40kb, respectively), in the size of the capsids, and in the fraction of free volume within the capsid filled by DNA. In addition, the interior of the P4 and phi29 capsids are empty, but T7 has an internal protein core that plays a role in the organization of the final packaged structure. We are preparing a paper for publication that describes the results of this work (Locker, Fuller & Harvey, *in preparation*).

Exploring Assembly Energetics of the 30S Ribosomal Subunit Using an Implicit Solvent Approach (Brooks, McCammon):

To explore the relationship between the assembly of the 30S ribosomal subunit and interactions among the constituent components, 16S RNA and proteins, relative binding free energies of the T. thermophilus 30S proteins to the 16S RNA were studied based on an implicit solvent model of electrostatic, nonpolar, and entropic contributions. The late binding proteins in our assembly map were found not to bind to the naked 16S RNA. The 5′ domain early kinetic class proteins, on average, carry the highest positive charge, get buried the most upon binding to 16S RNA, and show the most favorable binding. Some proteins (S10/S14, S6/S18, S13/S19) have more stabilizing interactions while binding as dimers. Our computed assembly map resembles that of E. coli; however, the central domain path is more similar to that of A. aeolicus, a hyperthermophilic bacteria.

Publications supported by the resource

Trylska, J, McCammon, JA & Brooks, CL, III. Exploring Assembly Energetics of the 30S Ribosomal Subunit using an Implicit Solvent Approach. *J Am Chem Soc* **in the press** (2005).

Protein Structure Prediction and Protein Folding

Protein modeling and structural genomics – the MMTSB Tool Set (Brooks):

The implementation of a modeling protocol for protein structure prediction involved several methodological advances that were developed within the MMTSB resource. We have devised an efficient novel procedure for accurate reconstruction of all-atom models from SICHO protein representations necessary for combining low- and high-resolution modeling and implemented this protocol as a component of the MMTSB Tool Set, a distributed community resource. Further efforts have focused on improving sampling efficiencies for simulations of both lattice models as well as all-atom models through the use of replica exchange techniques. In replica exchange simulations multiple structures are simulated using conventional Monte Carlo or molecular dynamics simulations in parallel at different, fixed temperatures and periodically exchanged according to Metropolis criteria based on their energy. This approach has shown to be extremely effective due to a combination of a simulated annealing protocol with the possibility of repeated refolding of poorly folded protein structures at higher temperatures until the lowest energy conformations, presumably at or near the native fold, are found. We have developed and tested computational tools for using this simulation technique successfully with lattice Monte Carlo simulations as well as all-atom molecular dynamics simulations in an implicit solvent environment. We participated in CASP6 and are currently enhancing the SICHO/MONSSTER scoring function to improve hydrogen bond and side chain interaction representations. Finally, work in beginning to incorporate the latest generation structure prediction tools, PROSPECTOR and TASSER into the MMTSB Tool Set and Predictor@home for distribution.

In continuing collaboration with Michael Feig (now an Assistant Professor ant Michigan State University), we are extending the analysis tool portion of the MMTSB Tool Set. Other additions including improvements of the replica exchange samplng approaches and the development of tools for NMR structure refinement.

Publications supported by the resource

M Feig, J Karanicolas, and CL Brooks, III, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modeling* **22** (2004) 377-95.

Predictor@home, protein structure prediction via the world-wide-web (Brooks):

This core research project is directed toward the development, implementation and application of structural bioinformatics methods to protein structure prediction, loop modeling and homology modeling. During the last year we have deployed a new computational infrastructure for protein structure prediction utilizing volunteer computing resources through a world-wide-web distributed computing mechanism. This architecture is based on the BOINC (Berkeley Open Infrastructure for Network Computing) middleware (http://boinc.berkeley.edu) and the software packages CHARMM (biomolecular simulation package) and mfold (lattice-based protein folding code, part of the MMTSB Tool Set) called predictor@home (http://predictor.scripps.edu). Predictor@home is a world-community experiment and effort to use distributed world-wide-web volunteer resources to assemble a supercomputer able to predict protein structure from protein sequence. Our work is aimed at testing and evaluating new algorithms and methods of protein structure prediction in the context of the Sixth Biannual CASP (Critical Assessment of Techniques for Protein Structure Prediction) experiment. The goal is to utilize these approaches together with the immense computer power that can be harnessed through the internet and volunteers all over the world to address critical biomedical questions of protein-

related diseases. Our ultimate objective in this project is to deploy these resources to provide a service component structure prediction server for the scientific community.

Publications supported by the resource

Taufer, M, An, C, Kerstens, A & Brooks, CL, III (2005) in *Fourth IEEE International Workshop on High Performance Computational Biology (HiCOMB 2005)*, Denver, Colorado), Vol. in the press.
Taufer, M, Anderson, D, Cicotti, P & Brooks, CL, III (2005) in *14th Heterogeneous Computing Workshop (HCW 2005)*, Denver, Colorado), Vol. in the press.

Multiresolution sampling methods for protein and peptide conformational space (Skolnick)

This core project focuses on the development, validation, and application of hybrid models designed to explore various aspects of protein and peptide folding, ligand docking and thermodynamics. In that regards, the recently developed TASSER algorithm was applied to predict the tertiary structures of all CASP6 targets. TASSER is a hierarchical approach that consists of template identification by the threading program PROSPECTOR_3, followed by tertiary structure assembly via rearranging continuous template fragments. For all 90 CASP targets/domains, PROSPECTOR_3 generates initial alignments with an average root-mean-square deviation to native (RMSD) of 8.4 Å with 79% coverage. After TASSER reassembly, the average RMSD decreases to 5.4 Å over the same aligned residues; the overall cumulative TM-score increases from 39.44 to 52.53. Despite significant improvements over the initial template alignment observed in all target categories, the overall quality of the final models is dictated by the threading templates quality. The average TM-score of TASSER models in the three categories are 0.79 (CM, 43 targets/domains), 0.47 (FR, 37 targets/domains), and 0.30 (NF, 10 targets/domains). This highlights the need to develop novel (or improved) approaches to identify very distant targets as well as better NF algorithms.

The ability of the combined AMBER/GBSA potential to identify native from misfolded structures as well as the correlation with native-likeness was assessed on a 30 protein test set, a subset of the PDB200 benchmark set of nonhomologous proteins representative of the PDB between 41 and 200 residues. For each protein, 14,000 decoys, previously generated by TASSER, were used. When the decoys were only locally minimized with AMBER/GBSA, nearly all of the native structures from the set are recognized as the lowest energy structures, with an average native-decoy energy gap of around 10% of the total native energy. The few failures involved proteins that were crystallized as a part of a complex. In a more difficult test, the decoys were subject to 100 ps Molecular Dynamics relaxation at 300° K followed by minimization. Now, the native structure is not the lowest energy one, and the AMBER/GBSA energy does not generally exhibit a correlation with native-likeness as assessed by the root-mean-square-deviation from native. In the coming year, the entire PDB200 set will be evaluated and an optimization of the AMBER/GBSA force field will be performed.

Publications supported by the resource

Kolinski, A & Skolnick, J. Reduced Models of Proteins and Their Applications. *Polymer* **45:** 511-24 (2004).
Bindewald, E & Skolnick, J. A scoring function for docking ligands to low-resolution protein structures. *J Comp Chem* **26:** 374-83 (2005).
Skolnick, J & Zhang, Y (2005) in *Systems Biology*, ed. Rigoutsos, I (Springer-Verlag, Vol. in the press.
Zhang, Y, Arakaki, A & Skolnick, J. TASSER: An automated method for the prediction of protein tertiary structures in CASP6. *Proteins* **in the press** (2005).

TAMD for efficient conformational sampling

We have recently implemented torsion angle molecular dynamics (TAMD) in CHARMM and investigated its application to augment conformational sampling of peptides and proteins [1]. Interesting conformational changes in proteins mainly involve torsional degrees of freedom. Carrying out molecular dynamics in torsion space does not only explicitly sample the most relevant degrees of freedom, but also allows larger integration time steps with elimination of the bond and angle degrees of freedom. However, the covalent geometry needs to be fixed during internal coordinate dynamics, which can introduce severe distortions to the underlying potential surface in the extensively parameterized modern Cartesian-based protein force fields. A ``projection" approach [2] is extended to construct an accurate internal coordinate force field (ICFF) from a source Cartesian force field. Torsion cross-term corrections constructed from local molecular fragments, together with softened van der Waals and electrostatic interactions, are used to recover the potential surface and incorporate implicit bond and angle flexibility. MD simulations of dipeptide models demonstrated that full flexibility in both the backbone ϕ/φ and side chain χ angles were virtually restored. The efficacy of TAMD in enhancing conformational sampling was then further examined by folding simulations of small peptides and refinement experiments of protein NMR structures. The results showed that an increase of several fold in conformational sampling efficiency can be reliably achieved. The current study also revealed some complicated intrinsic properties of internal coordinate dynamics, beyond energy conservation, that can limit the maximum size of the integration time step and thus the achievable gain in sampling efficiency.

Publications supported by the resource

Chen, J, Im, W & Brooks C. L., I. Torsion Angle Molecular Dynamics for Efficient Sampling of Protein Conformations. *J Comp Chem* **in the press** (2005).

Citations

[1] J. Chen, W. Im, and C.L. Brooks III. Application of Torsion Angle Molecular Dynamics for Efficient Sampling of Protein Conformation. *J. Comput. Chem.* in press (2005).
[2] V. Katritch, Totrov, M.M. and Abagyan, R.A. ICCF: A New Method to Incorporate Implicit Flexibility into an Internal Coordinate Force Field. *J. Comput. Chem.* 24(2): 254-65

Implicit solvation models for proteins, nucleic acids and membranes (Brooks):

This core research project is aimed at the development of implicit solvent models for representation of aqueous and membrane environments for the purposes of exploring the folding and modeling of (integral membrane) proteins and peptides. The focus has been in the exploration of generalized Born implicit solvent/implicit membrane models and extensions of the treatment of long-range dispersion interactions in these systems. Applications have been made and in progress to folding of proteins and peptides in aqueous solution, refinement of the structure in the presence of NMR derived restraints, the insertion, folding and structure determination of membrane bound peptides and proteins, and development of homology models

1. Membrane GB

Reformulating the calculation of the self-electrostatic solvation energy in GBSW (generalized Born model a simple smoothed switching function at the dielectric boundary) formalism [1], we were able to account for the heterogeneous influence of biological membranes in terms of a solvent-inaccessible infinite planar low-dielectric slab [2]. The reliability of the membrane GB model was assessed by exploring the role of biological membranes in influencing the conformational changes in membrane

(with melittin from bee venom), the tilt of transmembrane peptides with respect to the membrane normal (with the M2 protein from influenza A), and helix-helix interactions in membranes and the prediction of the configuration of transmembrane helical bundles (with the transmembrane domain of glycophorin A). By combining the membrane GB model with advanced computational sampling methods, like replica-exchange (REX) molecular dynamics (MD), we were able to fold and assemble helical membrane peptides correctly [2].

To further explore the applicability of the membrane GB model, we performed *de novo* folding simulations of the major pVIII coat protein from filamentous fd bacteriophage (fd coat protein) using REX-MD [3]. The quality of the predicted structures, judged by comparison of the root-mean-square deviations of a room temperature ensemble of conformations from the REX simulations and experimental structures from both solid-state NMR in lipid bilayers and solution-phase NMR on the protein in micelles, was quite good. Additional comparisons of computed solid-state NMR properties, the ^{15}N chemical shift and ^{15}N-^{1}H dipolar coupling constants, showed semi-quantitative agreement with the corresponding measurements.

The mechanism of interfacial folding and membrane insertion of designed peptides has been also explored using the membrane GB model and REX-MD [4]. Folding/insertion simulations initiated from fully extended peptide conformations in the aqueous phase, at least 28 Å away from the membrane interface, demonstrate a general mechanism for structure formation and insertion (when it occurs). The predominately hydrophobic peptides from the synthetic WALP and TMX series first become localized at the membrane--solvent interface where they form significant helical secondary structure via a helix--turn--helix motif that inserts the central hydrophobic residues into the membrane interior, and then fluctuations occur that provide a persistent helical structure throughout the peptide and it inserts with its N-terminal end moving across the membrane. More specifically, we observed that: (i) the WALP peptides (WALP16, WALP19, and WALP23) spontaneously insert in the membrane as just noted; (ii) TMX-1 also inserts spontaneously following a similar mechanism, and forms a transmembrane helix with a population of about 50% at 300 K; (iii) TMX-3 does not insert, but exists in a fluctuating membrane interface-bound form. These findings are in excellent agreement with available experimental data.

Recently, we have critically assessed the efficacy of the membrane GB model in predicting the assembled structure of three helical membrane peptides: transmembrane domains of glycophorin A (GpA: dimer), the M2 proton channel (M2-TMP: tetramer), and phospholamban (PLB: pentamer), where the native oligomerization states are given in parentheses [5]. The assembly REX-MD simulations of each peptide were carried out in various oligomerization states N (N=1, 2, 3, 4, 5, and 6) using the membrane GB model and imposing corresponding symmetries. We observed that: (i) the native-like assembled structures of each peptide were found to have the major populations in the room-temperature ensemble structures generated by REX-MD at the corresponding native oligomerization states and (ii) the calculated free energies including entropic contributions appeared to be minimum at N=5 for GpA, N=5 for M2-TMP, and N=4 for PLB. While the current model was able to provide free energy minima close to the native oligomerization of M2-TMP and PLB, close examination showed that the helix-helix van der Waals interactions play a dominant role in free energy calculations due to the lack of peptide-lipid van der Waals interactions, which is clearly problematic in GpA. The present studies indicate the limitation of the current model and a direction to its improvement toward *de novo* prediction of oligomerization states of membrane proteins and peptides.

The overall studies suggest the potential for the membrane GB model together with advanced simulation protocols to assist experimental approaches in exploring the nature and mechanism of membrane associated folding, insertion, and assembly of biologically important peptides.

2. NMR refinement

Electrostatic interactions are often oversimplified or ignored in the energy functions for NMR structure calculations, because it is difficult to evaluate them reliably without proper description of the dielectric screening by solvent. In light of recent improvements in implicit solvent models, it was showed that simulated annealing refinement in a GB implicit solvent could lead to noticeable improvement in the final protein NMR structures in terms of the backbone dihedral angle distributions and hydrogen bond patterns [6]. However, the impact of implicit solvent is rather small when a sufficient number of experimental restraints exist (such as in the final stage of NMR structure determination). In contrast, it was recently demonstrated that replica exchange molecular dynamics (REX-MD) refinement in a GB implicit solvent model could significantly improve the quality of structures and the radius of convergence when the experimental data is limited [7]. For example, while conventional structure calculations using an initial set of sparse NOE restraints were unable to identify a unique topology for a protein domain, high-quality native-like initial folds were generated through REX-MD refinement of the initial structures with a GB implicit solvent [8]. These models could be then used to make further assignments of ambiguous NOEs and speed up the structure determination process.

3. Optimized implicit solvent force fields

The successes and failures of various implicit solvent models in their applications to the biological problems arise in principle from their ability in maintaining delicate balance of energetics between sets of competing interactions, i.e., the solvation preference of each sidechain and the peptide backbone in aqueous bulk solution versus the strength of solvent-mediated interactions between these moieties in a complex protein environment [9]. To what extent a GB force field can capture this delicate balance appears to be a key in the success of its applications.

Recently, we have calibrated the backbone H-bond strength as well as water-mediated sidechain-sidechain interactions by adjusting the input radii for a GB model. Since the dielectric boundary in continuum dielectric solvent models dictates all of the electrostatic and most of nonpolar solvation energetics, it is physically appropriate to optimize the input radii not only based on solvation of individual sidechains but also in consideration of solvent-mediated interactions. Our study of conformational equilibria of two helical peptides, $(AAQAA)_3$ and α-lac, demonstrated that the helical content can be significantly influenced with small modification of the backbone input radii [9]. For instance, based on the explicit-solvent H-bond PMFs, a small change of the amide nitrogen input radius from 2.23 Å (the Nina's radii [10]) to 1.95 Å, which corresponds to 0.4 kcal/mol H-bond strength difference in an alanine dipeptide model, alters the average helicity of $(AAQAA)_3$ at 270 K from 83% to 55%, closer to the 50% value observed in experiments. In the case of α-lac, the same modification results in about a 25% reduction in the helicity of residues 108-111, which is in accord with the experimental results that this segment of the peptide is largely unstructured in water.

Based on the optimized implicit solvent force fields, we are presently examining the thermodynamic equilibria of various helical peptides and stability of various proteins, and making progress in folding of Trp-cage and Trp-zip mini proteins. The preliminary results are very encouraging and the manuscripts for publications are in preparation.

Publications supported by the resource

Chen, J, Im, W & Brooks, CL, III. Refinement of NMR structures using implicit solvent and advanced sampling techniques. *J Am Chem Soc* **126:** 16038-47 (2004).
Im, W & Brooks, CL, III. De novo folding of membrane proteins: an exploration of the structure and NMR properties of the fd coat protein. *J Mol Biol* **337:** 513-9 (2004).
Bu, L, Im, W & Brooks, CL, III. Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. *J Mol Biol* **submitted for publication** (2005).
Chen, J, Won, HS, Im, W, Dyson, HJ & Brooks, CL, III. Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *J Biomol NMR* **31:** 59-64 (2005).
Im, W & Brooks, CL, III. Chemical Theory and Computation Special Feature: Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc Natl Acad Sci U S A* **102:** 6771-6 (2005).
Im, W, Chen, J & Brooks, CL, III (2005).Peptide and protein folding and conformational equilibria: Theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models in *Peptide H-Bonds and Peptide Solvation*, eds. Baldwin, RE & Baker, DJ (Elsevier, Vol. in the press.

Citations

[1] W. Im, M.S. Lee, and C.L. Brooks III. Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* 24:1691-1702 (2003).
[2] W. Im, M. Feig, and C.L. Brooks III. An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys. J.* 85:2900-2918 (2003).
[3] W. Im and C.L. Brooks III. *De Novo* Folding of Membrane Proteins: An Exploration of the Structure and NMR Properties of the fd Coat Protein. *J. Mol. Biol.* **337**, 513-519 (2004).
[4] W. Im and C.L. Brooks III
Interfacial Folding and Membrane Insertion of Synthetic Peptides Studied by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. USA* 102:6771-6776 (2005).
[5] L. Bu, W. Im, and C.L. Brooks III
Membrane Assembly of Simple Helix Homo-oligomers Studied via Molecular Dynamics Simulations. *J. Mol. Biol.* submitted (2005).
[6] B. Xia, V.Tsui, D.A.Case, H.J.Dyson, and P.E.Wright. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR*, 22:317–331 (2002).
[7] J. Chen, W. Im, and C.L. Brooks III. Refinement of NMR Structure Using Implicit Solvent and Advanced Sampling Techniques. *J. Am. Chem. Soc.* 126:16038-16047 (2004).
[8] J. Chen, H. Won, W. Im, H.J. Dyson and C.L. Brooks III. Generation of Native-like Models from Limited NMR Data, Modern Force Fields and Advanced Conformational Sampling. *J. Biomol. NMR* 31:59-64 (2005).
[9] W. Im, J. Chen, and C.L. Brooks III. Peptide and Protein Folding and Conformational Equilibria: Theoretical Treatment of Electrostatics and Hydrogen Bonding with Implicit Solvent Models. *Adv. Protein Chem.* in press (2005)
[10] M. Nina, D. Beglov, and B. Roux. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. J. Phys. Chem. B, 101:5239–5248 (1997).

Multi-resolution modeling of molecular complexes with elastic network normal mode analysis

The structures and mechanics of myosin inhibition (Tama, Feig, Taylor, Brooks):

A combination of experimental structural data, homology modelling and elastic network normal mode analysis have been used to explore how coupled motions between the two myosin heads and the dimerization domain (S2) in smooth muscle myosin II determine the domain movements required to achieve the inhibited state of this ATP-dependent molecular motor. These physical models rationalize the empirical requirement for at least two heptads of noncoiled α-helix at the junction between the myosin heads and S2 and the dependence of regulation on S2 length. The results correlate well with biochemical data regarding altered conformational dependent solubility and stability. Structural models of the conformational transition between putative active states and the inhibited state show that torsional flexibility of the S2 α-helices is a key mechanical requirement for myosin II regulation. The results may have implications for function in other dimeric molecular motors.

Publications supported by the resource

Tama, F, Feig, M, Liu, J, Brooks, CL, III & Taylor, KA. The requirement for mechanical coupling between head and S2 domains in smooth muscle myosin ATPase regulation and its implications for dimeric motor function. *J Mol Biol* **345**: 837-54 (2005).

Virus swelling studied via elastic network normal modes (Tama, Brooks):

During the past five years there has been a tremendous growth of the number of virus structures solved by X-ray crystallography and cryo electron microscopy (EM) experiments, including virus capsids adopting several functionally relevant conformations. In addition, with time resolved small angle scattering experiments dynamics of viruses can now be monitored and for some systems several intermediate have been observed. These large-scale rearrangements are an integral part of the life cycle of the virus and have been observed in the maturation process, during assembly and may be involved in the release of RNA during infection. Disrupting the assembly or the maturation processes in viruses may be potential targets for antiviral activities. Thus, the description, prediction and exploration of such conformational changes for viruses are essential to understand their mechanism of activity and should be helpful in the development of new drugs.

Therefore, the mechanical properties and putative dynamical fluctuations of a variety of viral capsids comprising different sizes and quasi-equivalent symmetries were analyzed by performing normal mode analysis using the elastic network model. The expansion of the capsid to a swollen state is studied using normal modes and is compared with the experimentally observed conformational change for three of the viruses for which experimental data exists. A combination of one or two normal modes was shown to capture remarkably well the overall translation that dominates the motion between the two conformational states, and reproduces the overall conformational change. For all of the viral capsids it is observed that the nature of the modes is different. In particular for the T=7 virus, HK97, for which the shape of the capsid changes from spherical to faceted polyhedra, two modes are necessary to accomplish the conformational transition. In addition, extension of this study to viral capsids with other T numbers reveals the similarities and differences in the features of virus capsid conformational dynamics. In particular, the pentamers generally have higher flexibility and propensity to move freely from the other capsomers, which facilitates the shape adaptation that may be important in the viral life cycle.

Recently we have also started to explore the effect of RNA on the swelling process of viruses. We are currently studying the Satellite Tobacco Mosaic Virus (STMV) which is a T=1 icosahedral virus. About 60 % of its genomic material has been identified by X-ray crystallography. It has been observed experimentally that upon change in pH or RNA degradation, STMV adopts a swollen form (its radius increases by 5 Å). Using the lowest frequency - non-degenerate - normal mode to model the swelling process, strain energy analysis reveals that the swelling is stiffer when RNA is present which could potentially explain why STMV adopts a conformation with a larger radius when RNA is degraded.

Publications supported by the resource

Tama, F & Brooks, CL, III (2004).Unveiling molecular mechanisms of biological functions in large macromolecular assemblies using elastic network normal mode analysis in *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, eds. Cui, Q & Bahar, I (CRC Press.
Tama, F & Brooks, CL, III. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J Mol Biol* **345:** 299-314 (2005).

Structure refinement and modeling fitting into low resolution experimental data

NMFF – A new program for flexible fitting of atomic structures into EM maps (Tama, Brooks):

Cryo-EM is emerging as a primary tool not only for the elucidation of structural information but also for the examination of dynamical properties of supramolecular assemblies. To complement emerging information from EM, it is tempting to use existing X-ray data to construct atomic models. As large assemblies often undergo large functional rearrangements, the conformations of the available X-ray structures may not correspond to the conformation apparent in the cryo-EM data. In such cases, the fitting can be rather complicated, since not only the correct orientation needs to be found but conformational rearrangements must also be considered. However, current approaches deform the molecule in an *ad-hoc* way. Therefore, we have introduced a novel method for the quantitative flexible docking of a high-resolution structure into low-resolution maps of macromolecular complexes from EM that takes into account the conformational flexibility of biological systems.

The algorithm has been tested on several tests cases and we have demonstrated that the results are robust. We have also been working in close collaboration with several groups to fit high-resolution structures into experimental maps of variety of systems. The discussions with our collaborators, who will be the potential users of the algorithm, have been helpful to prepare a released version of NMFF. The release of NMFF for general use within the scientific community is planned for the summer 2005.

Publications supported by the resource

Tama, F, Miyashita, O & Brooks, CL, III. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* **337:** 985-99 (2004).

Tama, F, Miyashita, O & Brooks, CL, III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* **147:** 315-26 (2004).

Flexible refinement of atomic level structure into the GroEL-substrate complex (Tama, Brooks, Fisher, Gogel):

The 13Å resolution structures of GroEL bound to a single monomer of the protein substrate glutamine synthetase (GS_m), as well as that of unliganded GroEL have been determined from a heterogeneous image population using cryo-electron microscopy (cryo-EM) coupled with single particle image classification and reconstruction techniques. Structural data from cryo-EM maps and dynamic modeling were combined, taking advantage of the known X-ray crystallographic structure and normal mode flexible fitting (NMFF) analysis, to describe the changes that occur in GroEL structure induced by GS_m binding. The NMFF analysis reveals that the molecular movements induced by GS_m binding propagate throughout the GroEL structure. The modeled molecular motions show that some domains undergo en-bloc movements while others show more complex independent internal movements. Interestingly, the substrate-bound apical domains of both the *cis* (GS-bound ring) and *trans* (opposite substrate free ring) show counter-clockwise rotations, in the same direction (though not as dramatic) as those documented for the ATP-GroEL induced structure changes. The structural changes from the allosteric substrate protein-induced negative cooperativity between the GroEL rings involves upward concerted movements of both *cis* and *trans* equatorial domains toward the GS_m bound ring while the interring distances between the heptamer contact residues are maintained. Furthermore, the NMFF analysis also identifies the secondary structural elements that are involved in the observed ~5 Å reduction in the diameter of the cavity opening in the unbound *trans* ring. Understanding the molecular basis of these substrate protein-

induced structural changes across the heptamer rings provides insight into the origins of the allosteric negative cooperative effects that are transmitted over long distances (~140 Å).

Publications supported by the resource

Falke, S, Tama, F, Brooks, CL, III, Gogol, EP & Fisher, MT. The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J Mol Biol* **348:** 219-30 (2005).

Flexible refinement of atomic level structure into the LF-PA63h complex (Tama, Mitra, Brooks):

Anthrax is caused by one of the two virulence factors produced by the bacteria *Bacillus anthracis*. This factor is a cocktail of three monomeric proteins, lethal factor (LF, 90kD), edema factor (EF, 89kD) and the protective antigen (PA, 83kD).

Understanding the molecular recognition processes involved in the formation of the toxic Antrahx complex and the eventual delivery of the lethal and edema factors is critical to gaining insight into the pathogenesis of anthrax, and eventually to rational design of agents to block toxin action. Toward this end, we had directly visualized the PA63h.LF complex from analysis of images of specimens suspended in vitrified buffer by cryo-electron microscopy (Ren *et al.*, 2004). Comparison of the averaged 3-dimensional (3-D) views of the liganded and unliganded complex showed that the PA63h molecule binds to one LF molecule, which is localized to the non-membrane-interacting "top" face of the heptamer. The observed asymmetric binding of the ligand involves interaction with successive (at least 3) PA63 monomers and causes partial unravelling of the heptamer enlarging the lumen volume (Ren et al., 2004). This structural perturbation could be utilized to allow the ligand to ease into the vestibular lumen prior to its eventual translocation through the pore in the membrane.

The docking of one molecule of LF and the seven PA63 monomers in the density map of the complex was visually carried out, using the program O, based on the X-ray structures of PA63 and PA63h (Petosa *et al.*, 1997) and LF (Pannifer *et al.*, 2001) and using knowledge of the interacting regions of the two-polypeptide chains, as determined by site-directed mutagenesis studies. These, for PA63 have been identified (Cunningham *et al.* 2002) as well as for LF (Lacy *et al.* 2002). Based on the volume and shape of the density attributed to the LF in the cryo EM map generated by Ren *et al.* (2004), it was clear that LF has very likely undergone conformational rearrangement upon binding to the PA63 heptamer. In this study we have reanalyzed the model of the liganded complex using the normal mode flexible fitting (NMFF) procedure. In this technique, a linear combination of low-frequency normal modes, that are known to well represent conformational change in a molecule, is used in an iterative manner to deform the modeled structure optimally to conform to the low to medium resolution electron density map. Using NMFF and the previously determined 18Å resolution density map of the PA63h:LF complex (Ren *et al.*, 2004), a model for the ligand bound structure is constructed. The details of the heptamer structure, the observed conformational change of the LF molecule, and the nature of the ligand binding interface is discussed vis-à-vis their possible implication in the biological activity of anthrax toxin.

Citations

Cunningham, K., Lacy, D. B., Mogridge, J. and Collier, R.J. (2002). Mapping the lethal factor and edema factor binding sites on oligomeric anthrax protective antigen. Proc. Natl. Acad. Sci . USA. *99*, 7049-7053.

Lacy, D.B., Mourez, M., Fouassier, A. and Collier, R.J. (2002). Mapping the anthrax protective antigen binding site on the lethal and edema factors. J. Biol. Chem. *277*, 3006-3010
Pannifer, A.D. *et al.* (2001). Crystal structure of the anthrax lethal factor. Nature *414*, 229-233.
Petosa, C., Collier, R.J., Klimpel, K.R., Leppla, S.H. and Liddington, R.C. (1997). Crystal structure of the anthrax toxin protective antigen. Nature *385*, 833-838.
Ren , G., Quispe, J., Leppla, S. H. and Mitra, A. K. (2004) Large-scale structural changes accompany binding of lethal factor to anthrax protective antigen: a cryo-electron microscopic study. Structure *12*, 2059-2066

Publications supported by the resource

Tama, F, Ren, G, Leppla, SH, Brooks, CL, III & Mitra, AK. Model of the toxic complex of anthrax based on a cryo-electron microscopic investigation. *J Mol Biol* **to be submitted** (2005).

Ribosome structural studies (Tama, Frank, Brooks):

The proper functioning of any living cell requires that approximately half of all proteins synthesized in the cytosol must be translocated across (in the case of soluble proteins) or inserted into (in the case of membrane proteins) a cell membrane. Protein translocation at the membrane occurs through a proteinaceous channel, termed the translocon or translocase. The core of the translocon, the protein-conducting channel (PCC), is a heterotrimeric integral membrane protein complex (Sec**Y** in eubacteria/archaea, Sec**61** in eukaryotes) composed of the following: the α-subunit (SecY in eubacteria/archaea, Sec61α in mammals), the β-subunit (SecG in eubacteria, Secβ in archaea, Sec61β in mammals) and the γ-subunit (SecE in eubacteria/archaea, Sec61γ in mammals).

Cryo-EM reconstruction, performed by J. Frank and his collaborators, of the *E. coli* PCC, SecYEG, complexed with the ribosome and a signal sequence containing nascent chain, shows mRNA, three tRNAs, the nascent chain and detailed features of both an open, translocating PCC and a second, closed PCC bound to mRNA hairpins. The translocating PCC forms connections with ribosomal RNA hairpins on two sides and ribosomal proteins at the back, leaving a frontal opening. Normal mode flexible fitting was used to interpret the cryo-EM reconstruction in term of high-resolution structure. NMFF of the archaeal SecYEβ structure into the PCC EM densities favors a front-to-front arrangement of two SecYEG complexes in the PCC, and supports channel formation by the opening of two linked SecY halves during polypeptide translocation. Based on this observation in the open PCC of two segregated channels with different degrees of access to bulk lipid, a model for co-translational translocation was proposed

Publications supported by the resource

Mitra, K, Schaffitzel, C, Shaikh, T, Tama, F, Jenni, S, Brooks, CL, III, Ban, N & Frank, J. Structure of the E. coli protein-conducting channel bound to a translating ribosome. *Nature* submitted for publication (2005).

Red Clover Necrotic Mosaic Virus (Baker,Tama, Brooks):

Red clover necrotic mosaic virus is a small, icosahedral plant virus (*Tombusviridae* family) with a single-stranded RNA genome. RCNMV virions were analyzed by cryo-electron microscopy (cryoEM) and three-dimensional reconstruction at 8 Å resolution. The virion capsids have prominent protrusions; inside the capsid there is an inner cage formed by complexes of genomic RNA and amino terminal

portion of the coat protein. Separate reconstructions were obtained at 12 Å and 16 Å resolution, respectively, for virion samples that had Ca^{2+} alone or both Mg^{2+} and Ca^{2+} divalent cations selectively extracted. Virions treated in this way exhibited dramatic conformational changes, as obtained from NMFF, in both capsid and inner cage, though the particle diameters remained unchanged. Removal of Ca^{2+} ions alone appears to trigger movement and rotation of the S and P domains at the quasi 3-fold axis of symmetry. Additional removal of Mg^{2+} led to the formation of 11-13 Å diameter channels that extend through the virion shell at the quasi 3-fold axis of symmetry and noticeably alter the inner cage conformation. Despite their relatively small size, these channels are sufficiently large to allow RNA to pass through the protein capsid shell. Structural and infectivity studies suggest the divalent cation induced dynamics may be involved in viral reorganization, which renders the viral RNA accessible for translation early in infection.

Publications supported by the resource

Sherman, MB, Guenther, RH, Tama, F, Mikhailov, AM, Orlova, EV, Brooks, CL, III, Baker, TS & Lommel, SA. Tentative mechanism of RCNMV replication. *J Mol Biol* **to be submitted** (2005).

Outreach, services and web sites and workshops and training

Current status of the MMTSB web site and related resources (Brooks):

The MMTSB web site (http://mmtsb.scripps.edu) is maintained to provide web services to the scientific community, distribute and support software tools developed within the resource, offer general information about the MMTSB resource and its projects, and provide selected links to related resources. Already available web services include the very successful and established Virus Particle Explorer (VIPER) as well as some utility functions for nucleic acid and protein modeling applications, in particular for interconversion between reduced lattice models and all-atom representation and the Go-model server. More recently we have released the Predictor@home web site (http://predictor.scripps.edu) for volunteer distributed computing protein structure prediction. We are preparing additional web services for protein structure evaluation, structure refinement, and limited protein structure prediction with a focus on loop modeling as the development and testing of the necessary modeling tools nears completion and additional computational resources needed for this kind of service are coming online. Several software packages for modeling applications are supported by the resource with links on the MMTSB web site to the corresponding distribution and documentation sites. They include CHARMM and Amber for general modeling applications, MONSSTER for lattice-based protein modeling, NAB, Yammp and YUP for nucleic acid modeling as well as Situs and XtalView for fitting models to X-ray and cryoEM maps. In addition, a new software package (NMFF) for the flexible refinement of high-resolution X-ray structures into low-resolution data from cryo-electron microscopy is being developed for distribution and service related refinement projects.

The MMTSB web site itself is the distribution site for the MMTSB Tool Set that is being developed for multiscale protein modeling and structure prediction application. It is a collection of modules, mostly written in perl to ensure portability and allow for easy modifications and extensions. This tool set provides interfaces to CHARMM, Amber, and MONSSTER specifically geared towards protein modeling applications and implements new methods developed within the resources for model interconversion and advanced sampling techniques. The design of the tool set is highly modular offering easy-to-use application-oriented utilities as well as flexible programming interfaces for more complex customized applications. Although continuing development, the current version of the MMTSB tool set is quite complete with respect to the available methods. Recently continue to expand its functionality towards more comprehensive support of the Amber molecular modeling package. Motivated by newly developed generalized Born models for implicit solvation that have become in Amber, we have added the ability to run molecular dynamics simulations and enhanced sampling replica exchange simulations with Amber. We have also introduced new multiscale sampling protocols that take advantage of the unique architecture of the tool set. They involve Monte Carlo type simulations with different types of moves that consist of short lattice simulations followed by energy evaluations on an all-atom energy surface. This protocol is meant to provide both efficient sampling and an accurate energy function for applications such as structure refinement. Other changes involve minor additions, optimized parameters, and bug fixes. In the future, we are planning to extend and test the newly introduced multiscale sampling protocols and offer an alternative graphics-based user interface.

The full version has been available for download on the MMTSB web site for more than two years and is thoroughly tested by now. Since March 2001 we have registered more than 500 unique downloads, most of them from academic research groups in the U.S. and abroad in related areas of research. Many of the users have and continue to provide valuable feedback for making the tool set more robust and add new functionalities. Our user base has been continuously increasing, and as expected it increased as a result of our first MMTSB workshop held in collaboration with the Pittsburgh Supercomputer Center in

June of 2003, our participation in the workshop "A Practical Course in Molecular Microscopy", November 12-20, 2003 at the NCRR supported Center for Integrative Molecular Biosciences (CIMBio) and the publication of a paper describing the design and applications of the MMTSB Tool Set. We anticipate this growth to continue as we

In order to provide adequate support extensive documentation of all tools and packages is available on the web site. In addition a tutorial focusing on specific applications is being developed and further support is available through email.

Publications supported by the resource

J Karanicolas and CL Brooks, III, An evolution of minimalist models for protein folding: from the behavior of protein-like polymers to protein function. *Biosilico* **2** (2004) 127-33.

Modeling Virus Assembly via Structure, Energy and Thermodynamics and the ViPEr web site (Reddy, Johnson, Brooks):

This project is directed towards the specific aims: 1) develop and maintain the VIrus Particle ExploreR (VIPER), a database and website, http://mmtsb.scripps.edu/viper/, of icosahedral virus structures 2) devise the structural and computational tools to analyze virus structures.

In a significant development in the past year, we successfully ported the entire VIPER site, all the data (coordinates), related information (e.g., taxonomical) and computationally derived results, into a MySQL relational database environment. This new database, which we now call VIPERdb, is currently accessible at the URL http://viperdb.scripps.edu will replace existing VIPER site in the year 2006. Furthermore, VIPERdb contains a lot more entries (211) of capsids that include all the viral capsid entries: virus-drug complexes, mutant structures and models derived based on cryoEM reconstructions. The database was created by Dr. Craig Shepherd, as postdoctoral fellow with assistance from Mr. Ian Borelli under the supervision of Dr. Vijay Reddy.

The derived data and pictorial (graphical) descriptions for all the entries were regenerated. The derived results include, contact tables, subunit-subunit buried surface areas, solvation energies, association energies, measures of quasi-equivalence (Q-scores) and accessible surface profiles and are stored as tables in the database. Some of these properties can be plotted on graphs as a function of residue numbers using JpGraph and PHP scripts in real time in a browser window. The location of the individual contacts in the contact tables can be visualized in real time on a browser widow and interacted with using VRML plugins.

To enable the users to gain access to the VIPERdb, a relational database, a window to browse the database is provided on the website to query the database. For instance the following query submitted through such window indentifies the highly exposed cysteine residues in a particular virus (Cowpea Mosaic Virus, CpMV) and sort in the descending order based on their radial distance.

query: SELECT viper.entry_id, label_asym_id, label_comp_id, label_seq_id, eff_rad_x_sasa_min, sasa_bound, radius_aa FROM viper_residue_asa, layer, viper WHERE viper_residue_asa.entry_id =layer.entry_id AND viper_residue_asa.entry_id = viper.entry_id AND sasa_bound > 0 AND viper.entry_id = '1ny7' AND (label_comp_id ='CYS')order by radius_aa desc

ENTRY_ID	LABEL_ASYM_ID	LABEL_COMP_ID	LABEL_SEQ_ID	SASA_BOUND	RADIUS_AA

1ny7	A	CYS	4	21.4	128.8
1ny7	C	CYS	119	8.4	117.8
1ny7	B	CYS	295	20.2	112.5
1ny7	C	CYS	132	2.7	112.3
1ny7	B	CYS	355	34.7	108.5
1ny7	C	CYS	177	1.5	106.7
1ny7	B	CYS	187	85.9	105.9

The power and utility of such a database is immense for comparing residue properties across different viruses within a family and even across families. Such comparisons are currently on going

We are extremely happy to report that VIPERdb was featured in the netwatch column in 18 Feb. 2005 issue of science magazine.

Another front of significant development in the past year is EMDB, a website (http://mmtsb1.scripps.edu/viper/em_index.php) of cryo-EM structures (densities) of icosahedral viruses. The EMDB is being developed by Dr. Padmaja Natarajan under the supervision of Professor Jack Johnson as extension and augment the VIPER database. There are currently over 40 EM reconstructions for which both electron densities and structure factor amplitudes and phases are available through this web site. A review article on how to use VIPER site accepted for publication in Nature reviews in Microbiology.

Publications supported by the resource

Natarajan, P, Lander, G, Shepherd, CM, Reddy, VS, Brooks, CL, III & Johnson, JE. Exploring Icosahedral Virus Structures with VIPER. *Nature Rev Micro Biol* **in the press** (2005).

Analysis of protein-protein interactions and quasi-equivalence in icosahedral virus capsids (Reddy, Brooks):

Viral capsids are composed of multiples copies of one or a few coat protein subunits. Knowledge on the extent of interactions of a capsid protein subunit in native and sometimes in non-native capsid environments will be of importance in understanding the relationship between the protein building block and the capsid size and architecture. Furthermore, the extent of similarity in the bonding and interactions of the subunits would point to the quasi-equivalences of subunits in viral capsids. The knowledge on the relationship between the type of the protein and the resulting capsid would aid in designing new and novel capsids.

The objective of this project is to come up with a measure/property that would quantify the extent to which a subunit is involved in the protein-protein interactions. To this end, we came up with fraction buried surface area of the subunit, which we call protein-protein interaction (PPI) index, as a measure of extent of protein-protein interactions correspond to a subunit in the capsid environment.

The solvent accessible surface areas (SASA) were calculated using the CHARMM molecular mechanics software package, using the CHARMM19 set of united atom parameters and a solvent probe radius of 1.4 Å. Buried surface area (BSA) was calculated by subtracting the SASA values of a subunit in the context of the capsid from the SASA of the same subunit in isolation. The SASA of the reference subunit in the context of the capsid was determined using a CHARMM stream file (script), which loops over and subtracts the BSAs due to every nearest subunit from the SASA of the isolated reference

subunit. For capsids with **T>1** the average values of protein-protein interaction (PPI) indices of the distinct subunits in each capsid are plotted with the standard deviations are shown as error bars.

In the case of T=1 capsids, the PPI indicies show an increase proportional to the molecular weight (size) of the coat protein subunit, where as the indicies for the subunits of the quasi-equivalent capsids interestingly are rather independent of the size. Remarkable outcome of such a study was the distinct clustering of "T=2" capsids with low PPI indicies and larger subunit molecular weights, which uniquely distinguish them as the thin shelled capsids with lager surface area per unit molecular weight. Similar analyses done as a function of capsid diameter also result in analogous results.

The variation (residual) in PPI indices of subunits of quasi-equivalent capsids was used as a measure to estimate extent of quasi-equivalence (1-residual) in capsids with T number > 1. This study has been published in the proteins in the year 2005.

Publications supported by the resource

Shepherd, CM & Reddy, VS. Extent of protein-protein interactions and quasi-equivalence in viral capsids. *Proteins* **58:** 472-7 (2005).

Developing tools for the analysis of virus structures – The ViPEr Tool Set (Reddy, Brooks):

The VIPER tool set contains a number analysis tools that have been developed over the past few years to analyze virus structures particularly in an automatic fashion, which also can be employed for the analysis of multimeric proteins. The objective of this project is to organize the VIPER toolset for distribution on the same lines as the MMTSB Tool Set.

These structural tools (Table 1) have been modified to generate the derived results/data to be uploaded to into a VIPERdb. In a significant addition to the tools we developed a new an algorithm and tool that automatically generates the PDB_to_VIPER matrix, which transforms the .pdb coordinates into VIPER convention. This tool is a hugh step forward in removing the bottleneck in performing the VIPER analysis in an automatic fashion.

All the resultant analyses from these tools can be done currently through the web interfaces from the VIPERdb site. Efforts are on going to generate a standalone tools set for the distribution to the structural virology community.

In addition a number of web-based utilities such as oligomer generator, gallery maker, contact finder, etc., which now operate by interacting with the relational database are also available at the VIPERdb (http://viperdb.scripps.edu) site.

Table 1. Various analysis tools available currently as part of the VIPER toolset.

S.No.	Name of the utility*	Function of the utility
1	process_pdb.pl	reads in pdb files and outputs CHARMM coordinate files
2	gen_hydr.stream	generates hydrogens and outputs .chr files
3	gen_segments.stream	Generates duplicate chains of subunits in the icos. asymmetric unit.
4	list_contacts.stream	Given NCS matrices & chains generates a list of possible subunit pairs
5	select_contacts.stream	Identifies a subset of the subunits pairs that are truly in contact
6	rotate_generic.stream	Transforms the coordinates by a chosen NCS matrix

7	coor_dist.stream	Identifies the number of residues pairs in contact, cutoff limit 8.5A
8	count.pl	Counts the number of residue pairs in contact that have been identified
9	select_unique.pl	Selects the unique subunit pairs that are in contact.
10*	gen_contact_table.stream	identifies and organizes residue pairs in contact at the subunit interfaces into columns
11*	qscore.stream	Calculate similarity score based on common residue pairs at the subunit interfaces
12*	bsa_solv_energy.stream	Calculates the solvation energies of subunit pairs based on buried ASAs.
13*	ppi.stream	Calculates protein-protein interaction index
14*	select_insert_site.stream	Identifies the potential sites for the insertion of loops on the viral surface.
15*	net_charge.stream	Caclculates net charge on the virus surface.
16*	pdbtoviper.c	Generates PDB to VIPER matrix

Files with ".pl" extension are perl scripts and ".stream" are CHARMM stream (script) files. Utilities identified by the "*" refers to the tools that calculates various derived properties/results.

Rewriting YAMMP (Harvey, Tang):

The rewriting of Yammp into YUP (Yammp Under Python) is complete, and YUP has been released. It is available from our website (http://rumour.biology.gatech.edu/YammpWeb/).
The rewrite achieved two long-standing goals: (1) integration of a scripting language and (2) the ability to assemble the force field for all types of models and especially non-standard ones. In the new version of the program, energy and force calculations form the computational core, and these functions remain written in the C programming language and compiled into a number of shared libraries of native machine code. Molecular mechanics methods such as energy minimization and molecular dynamics (MD) are implemented as Python programs that call the shared code.
The past year has seen two major advances.
First, we developed a prototype graphical user interface for Yammp (YupSee). It has two major components, one for setting up, carrying out, and analyzing simulations, and the other for displaying the resulting models. We have decided that the further development of YupSee will require an effort and resources that are beyond the scope of this grant, so we are seeking separate funding for this. (An R01 proposal was submitted on June 24, 2005 for this purpose, in response to PAR-03-106.)

Second, we have begun major efforts at Yammp dissemination. A series of five tutorials has been developed and posted online (http://rumour.biology.gatech.edu/YammpWeb/). The goal is to develop these to the point where they are stand-alone exercises, so that new users can download Yammp and the tutorials from the web, work their way through the tutorials, and begin using Yammp without assistance from anyone at Georgia Tech. Before this can be done, however, we are running a series of workshops to train new users, and to gain feedback for revising the tutorials. The first YUP training workshop was run at Georgia Tech from June 28 – July 1, 2005, and the second was run July 5-8. Each of these had five participants. As a result of user feedback, we have expanded the first tutorial (which teaches the use of Python) and are requiring future trainees to complete this before coming to the workshops. This enables us to shorten the workshop itself to three days. A third workshop is scheduled for August 16-18, and four people are currently enrolled.

Dissemination: Publications acknowledging NCRR from MMTSB Researchers in 2003-2004:

1. Beveridge, DL, Barreiro, G, Byun, KS, Case, DA, Cheatham, TE, 3rd, Dixit, SB, Giudice, E, Lankas, F, Lavery, R, Maddocks, JH, Osman, R, Seibert, E, Sklenar, H, Stoll, G, Thayer, KM, Varnai, P & Young, MA. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* **87:** 3799-813 (2004).
2. Chen, J, Im, W & Brooks, CL, III. Refinement of NMR structures using implicit solvent and advanced sampling techniques. *J Am Chem Soc* **126:** 16038-47 (2004).
3. Feig, M, Onufriev, A, Lee, MS, Im, W, Case, DA & Brooks, CL, III. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comp Chem* **25:** 265-84 (2004).
4. Feig, M, Karanicolas, J & Brooks, CL, III. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model* **22:** 377-95 (2004).
5. Feig, M & Brooks, CL, III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* **14:** 217-24 (2004).
6. Feig, M, Im, W & Brooks, CL, III. Implicit solvation based on generalized Born theory in different dielectric environments. *J Chem Phys* **120:** 903-11 (2004).
7. Ferrara, P, Gohlke, H, Price, DJ, Klebe, G & Brooks, CL, III. Assessing scoring functions for protein-ligand interactions. *J Med Chem* **47:** 3032-47 (2004).
8. Gohlke, H, Kuhn, LA & Case, DA. Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* **56:** 322-37 (2004).
9. Gohlke, H & Case, DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comp Chem* **25:** 238-50 (2004).
10. Im, W & Brooks, CL, III. De novo folding of membrane proteins: an exploration of the structure and NMR properties of the fd coat protein. *J Mol Biol* **337:** 513-9 (2004).
11. Karanicolas, J & Brooks, CL, III. An evolution of minimalist models for protein folding: from the behavior of protein-like polymers to protein function. *Biosilico* **2:** 127-33 (2004).
12. Kolinski, A & Skolnick, J. Reduced Models of Proteins and Their Applications. *Polymer* **45:** 511-24 (2004).
13. Kondo, D, Taufer, M, Brooks, CL, III, Casanova, H & Chien, A (2004) in *Proceedings of IPDPS 2004, IEEE/ACM International Parallel and Distributed Processing Symposium*, Santa Fe, New Mexico).
14. Konecny, R, Trylska, J, Tama, F, Zhang, D, Baker, NA, Brooks, CL, III & McCammon, JA. Electrostatic Properties of Cowpea Chlorotic Mottle Virus and Cucumber Mosaic Virus Capsids. *Proteins* **submitted for publication** (2004).
15. LaMarque, JC, Le, TV & Harvey, SC. Packaging double-helical DNA into viral capsids. *Biopolymers* **73:** 348-55 (2004).
16. Mackerell, AD, Jr., Feig, M & Brooks, CL, III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comp Chem* **25:** 1400-15 (2004).
17. MacKerell, AD, Jr., Feig, M & Brooks, CL, III. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc* **126:** 698-9 (2004).
18. Natrajan, A, Crowley, M, Wilkins-Diehr, N, Humphrey, MA, Fox, AD, Grimshaw, AS & Brooks, CL, III. Studying protein folding on the Grid: experiences using CHARMM on NPACI resources under Legion. *Concurrency and Computation: Practice and Experience* **16:** 385-97 (2004).
19. Rao, ALN & Reddy, VS (2004).Architecture of plant viruses. in *Handbook of Plant Virology*, eds. Khan, JD & Dijkstra, J (Haworth Press.

20. Reddy, VS & Johnson, JE. Structure derived insights into virus assembly. *Adv Virus Res* **in the press** (2004).
21. Reddy, VS, Natarajan, P, Lander, G, Qu, C, Brooks, CL, III & Johnson, JE (2004).Virus Particle Explorer (VIPER): A repository of virus capsid structures in *Conformational proteomics of macromolecular Architecture*, eds. Cheng, RH & Hammer, L (World Scientific Pub Co., Singapore), Vol. Chpt. 20, pp. 403-12.
22. Roberts, VA, Case, DA & Tsui, V. Predicting interactions of winged-helix transcription factors with DNA. *Proteins* **57:** 172-87 (2004).
23. Tama, F, Miyashita, O & Brooks, CL, III. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* **337:** 985-99 (2004).
24. Tama, F & Brooks, CL, III (2004).Unveiling molecular mechanisms of biological functions in large macromolecular assemblies using elastic network normal mode analysis in *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, eds. Cui, Q & Bahar, I (CRC Press.
25. Tama, F, Miyashita, O & Brooks, CL, III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* **147:** 315-26 (2004).
26. Taufer, M, Crowley, M, Price, D, Chien, A & Brooks, CL, III. Study of an Accurate and Fast Protein-Ligand Docking Algorithm based on Molecular Dynamics. *Concurrency and Computation: Practice and Experience* **in the press** (2004).
27. Tihova, M, Dryden, KA, Le, TV, Harvey, SC, Johnson, JE, Yeager, M & Schneemann, A. Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *J Virol* **78:** 2897-905 (2004).
28. Trylska, J, Konecny, R, Tama, F, Brooks, CL, III & McCammon, JA. Ribosome motions modulate electrostatic properties. *Biopolymers* **74:** 423-31 (2004).
29. Wriggers, W, Chacon, P, Kovacs, JA, Tama, F & Birmanns, S. Topology representing neural networks reconcile biomolecular shape, structure, and dynamics. *Neurocomputing* **56:** 365-79 (2004).
30. Bindewald, E & Skolnick, J. A scoring function for docking ligands to low-resolution protein structures. *J Comp Chem* **26:** 374-83 (2005).
31. Bu, L, Im, W & Brooks, CL, III. Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. *J Mol Biol* **submitted for publication** (2005).
32. Case, DA, Cheatham, T, Darden, T, Gohlke, H, Luo, R, Merz, KM, Jr., Onufriev, A, Simmerling, C, Wang, B & Woods, R. The Amber Biomolecular Simulation Programs. *J Comp Chem* **in the press** (2005).
33. Chen, J, Im, W & Brooks C. L., I. Torsion Angle Molecular Dynamics for Efficient Sampling of Protein Conformations. *J Comp Chem* **in the press** (2005).
34. Chen, J, Won, HS, Im, W, Dyson, HJ & Brooks, CL, III. Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *J Biomol NMR* **31:** 59-64 (2005).
35. Dixit, SB, Beveridge, DL, Case, DA, T.E. Cheatham, I, Giudice, E, Lankas, F, Lavery, R, Maddocks, JH, Osman, R, Sklenar, H, Thayer, KM & Varmai, P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II. Sequence contect effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys J* (2005).
36. Falke, S, Tama, F, Brooks, CL, III, Gogol, EP & Fisher, MT. The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J Mol Biol* **348:** 219-30 (2005).
37. Guvench, O & Brooks, CL, III. Folding Trp-cage to atomic resolution: Searching parameter space for a free energy minimum. *J Am Chem Soc* **submitted for publication** (2005).

38. Im, W & Brooks, CL, III. Chemical Theory and Computation Special Feature: Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc Natl Acad Sci U S A* **102:** 6771-6 (2005).
39. Im, W, Chen, J & Brooks, CL, III (2005).Peptide and protein folding and conformational equilibria: Theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models in *Peptide H-Bonds and Peptide Solvation*, eds. Baldwin, RE & Baker, DJ (Elsevier, Vol. in the press.
40. Mears, JA, Stagg, SM & Harvey, SC (2005).Modeling of Very Large RNA Assemblies Using Reduced Representations in *Handbook of RNA Biochemistry*, eds. Westhof, E, Binereif, S, Schön, A & Hartmann, R (Wiley-VCH, Weinheim), Vol. in the press.
41. Mitra, K, Schaffitzel, C, Shaikh, T, Tama, F, Jenni, S, Brooks, CL, III, Ban, N & Frank, J. Structure of the E. coli protein-conducting channel bound to a translating ribosome. *Nature* **submitted for publication** (2005).
42. Natarajan, P, Lander, G, Shepherd, CM, Reddy, VS, Brooks, CL, III & Johnson, JE. Exploring Icosahedral Virus Structures with VIPER. *Nature Rev Micro Biol* **in the press** (2005).
43. Shepherd, CM & Reddy, VS. Extent of protein-protein interactions and quasi-equivalence in viral capsids. *Proteins* **58:** 472-7 (2005).
44. Sherman, MB, Guenther, RH, Tama, F, Mikhailov, AM, Orlova, EV, Brooks, CL, III, Baker, TS & Lommel, SA. Tentative mechanism of RCNMV replication. *J Mol Biol* **to be submitted** (2005).
45. Skolnick, J & Zhang, Y (2005) in *Systems Biology*, ed. Rigoutsos, I (Springer-Verlag, Vol. in the press.
46. Tama, F & Brooks, CL, III. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J Mol Biol* **345:** 299-314 (2005).
47. Tama, F, Feig, M, Liu, J, Brooks, CL, III & Taylor, KA. The requirement for mechanical coupling between head and S2 domains in smooth muscle myosin ATPase regulation and its implications for dimeric motor function. *J Mol Biol* **345:** 837-54 (2005).
48. Tama, F, Ren, G, Leppla, SH, Brooks, CL, III & Mitra, AK. Model of the toxic complex of anthrax based on a cryo-electron microscopic investigation. *J Mol Biol* **to be submitted** (2005).
49. Taufer, M, An, C, Kerstens, A & Brooks, CL, III (2005) in *Fourth IEEE International Workshop on High Performance Computational Biology (HiCOMB 2005)*, Denver, Colorado), Vol. in the press.
50. Taufer, M, Anderson, D, Cicotti, P & Brooks, CL, III (2005) in *14th Heterogeneous Computing Workshop (HCW 2005)*, Denver, Colorado), Vol. in the press.
51. Trylska, J, McCammon, JA & Brooks, CL, III. Exploring Assembly Energetics of the 30S Ribosomal Subunit using an Implicit Solvent Approach. *J Am Chem Soc* **in the press** (2005).
52. Zhang, Y, Arakaki, A & Skolnick, J. TASSER: An automated method for the prediction of protein tertiary structures in CASP6. *Proteins* **in the press** (2005).

HIGHLIGHTS

Multi-resolution sampling for protein folding

Significant progress continues to be made in integrating reduced and full atomic protein models and a pipeline for protein structure prediction. This year included crucial tests of the infrastructure we have built and distribute for protein structure prediction with our participation in the 6th Annual CASP experiment utilizing the MMTSB Tool Set (MONSSTER) structure prediction pipeline with Predictor@home (our world-wide-web volunteer computing structure prediction supercomputer based on BOINC) and the PROSPECTOR/TASSER structure prediction pipeline. Performance by both the TSRI-based and Buffalo-based groups was good.

Exploring global motions in biological machines

A significant advance within our resource has been the development and testing of elastic normal mode models to explore large-scale displacements and rearrangements of biological assemblies associated such as virus capsids, the ribosome and myosin. During the current year we explored the similarities in "breathing motions" across viral capsids of vastly varied complexity and discovered general features shared by capsids displaying T=1 to T=13 architectures, providing new connections to common functional aspects of these systems. Additionally, working with resource collaborator, K. Taylor, from Florida State University, we continued out modeling of myosin dynamics, focusing most recently on myosin V. Interactions with Tom Smith from the Danforth Plant Science Institute to explore the cooperative motions in Glutamate Dehydrogenase have also been productive.

Development and testing of continuum solvent models to study electrostatic interactions in large systems and integral membrane peptides and proteins

A key goal of the resource is the development of algorithms that will allow rapid evaluation of conformational energies for large systems. We have enjoyed fantastic progress in this area during the past year. The development and testing of new generalized Born theories to treat the membrane implicitly have opened the door to studies of the structure and folding, as well as insertion and assembly of integral membrane proteins. These techniques continue to be fully integrated with our modeling approaches and re-distributed to the community through the CHARMM and Amber programs. Also of note in this area has been the application of the parallel Poisson-Boltzmann solvers developed by McCammon and his group to explore the electrostatic consequences of large-scale structural fluctuations in viruses and the ribosome and the integration of these methods with CHARMM. Finally, we continue to develop an MMTSB-based interface for protein structure determination via NMR utilizing our generalized Born models, replica-exchange sampling and NMR derived distance restraints.

Outreach, workshops and dissemination and service

During the past year we have participated in workshops on genome-scale modeling using the MMTSB Tool Set with Amber and CHARMM as well as our inaugural workshops on YUP/yammp. Each of these workshops was well received and. In November we will be involved in our second EM data fitting and model building wrkshop (in collaboration with the RR Center NRAMM headed by Bridget Carragher). It was during our first efforts with this workshop that a number of fruitful and important collaborations were initiated with Tom Smith, Alok Mitra, Mark Fisher and Joachim Frank. In addition, we have hosted visitors from a number of laboratories during the past year. For example, Ken Taylor, from Florida State University, returned for his second month-long stay in the Center and focused his efforts on using the Tool Set and our elastic network normal mode models to develop structural models for functionally important states of myosin that had been characterized via low-resolution electron microscopy. Also, recurring visits by Harold Scheraga (from Cornell University) and Mark Olson (from

the Army Laboratory for Biological Research, Fredrick, MD) invigorate our efforts in protein folding and loop modeling. Finally, the ViPER web-site and ViperDB have become the de facto source for virologists and structural biologists exploring these systems. Recent work with Chandra Bajaj and Tom Ferrin are brining new visualization tools to this enterprise..

DNA and genome packaging in viruses

Investigating the structures, kinetics and thermodynamics of the packaging of nucleic acids into viruses is of key importance in understanding viral life-cycles. Molecular mechanics approaches have been used to successfully mimic the packaging of double-helical DNA into model bacteriophage spherical capsids, compressing the in vivo timescale (10-1000 seconds) into the microsecond timescale. The resulting structures are offer new insights into the structural basis of packaging for a wide variety of experimental observations.

ORGANIZATION OF THE MMTSB RESOURCE

Professor Charles L. Brooks III directs the NCRR supported Center for the development of Multi-scale Modeling Tools for Structural Biology and in this capacity oversees the activities of the Center. He is a faculty member of the Department of Molecular Biology, with Peter Wright serving as Chair of this department. On matters related to the scientific and administrative aspects of the Center, Professor Brooks interacts with Peter Wright and Jeff Kelly, Vice President for Academic Affairs.

Professor Brooks and the MMTSB scientists are advised by an Advisory Committee comprised of scientists with expertise covering the areas of research and technology addressed in the Center. The Advisory Committee communicates on an annual basis with the Center scientists through an Advisory Committee Annual meeting and through more frequent contacts between committee members and Brooks. The members of the current Advisory Committee include:

- Professor Carol B. Post, Department of Medicinal Chemistry, Purdue University (Chair)
- Professor Igancio Tinoco, Department of Chemistry, University of California, Berkeley
- Processor J. Andrew McCammon, Department of Chemistry and Biochemistry, Howard Hughs Medical Institute, University of California, San Diego
- Professor Tim Baker, Department of Biological Sciences, Purdue University
- Professor Wah Chiu, Department of Biochemistry and Molecular Biology, Baylor College of Medicine
- Professor L. Ridgway Scott, Department of Computer Science and Department of Mathematics, University of Chicago

ADVISORY COMMITTEE REPORT

Our advisory board meeting is scheduled for September 21, 2005. We will forward the report once completed. Included on the Advisory Committee report page below is our most recent report from the September, 2004 meeting.

ALLOCATION OF RESOURCE ACCESS (ADVISORY COMMITTEES)

At this point in the development of our resource, there are no allocable resources. Thus, we have no allocation advisory committees.

DISSEMINATION OF INFORMATION

We continue to disseminate information regarding activities of the MMTSB resource through four main mechanisms. (i) The establishment of a WWW sites with resource and service related items (http://mmtsb.scripps.edu and http://mmtsb.scripps.edu/viper.html) (ii) Organizing and running training and research workshops here at TSRI and elsewhere. (iii) A visitors program has brought prominent scientists (Mark Olson, Harold Scheraga, Ken Taylor, Alex MacKerrel and students such as Sagar kathuria from Bob Matthews group at U. Mass, Medical School) to visit Scripps and learn about our resource activities; (iv) continued development and distribution of the software packages Amber, CHARMM, NAB, YAMMP (YUP), NMFF and the MMTSB toolset to the scientific community.

The ViPEr website has become the de facto standard used by the virology community. We have recently solicited feedback from ViPEr users regarding papers published that reference the ViPeR site. Below are a list of some of the papers which make reference to this site.

TRAINING

During the past year we have taught four workshops in the area of multi-scale modeling and will be involved in one in the area of electron microscopy and modeling atomic models into EM density maps held in collaboration with the NRAM (an NCRR supported resource directed by Bridget Carragher and Clint Potter) in November, 2005. The lecture schedule for this workshop can be found at http://nramm.scripps.edu/seminars/2005/cryoem/.

Most recently, the MMTSB program teamed up with the Pittsburgh Supercomputer Center to present a workshop on "Biomolecular Simulations Using VMD, Amber and MMTSB Toolkit." This workshop was held June 9-12, 2005 in Pittsburgh, and had 33 participants (chosen from over 90 applicants, a record number for this workshop program that has been ongoing for more than a decade.). The workshop covered basic to moderately advanced biomolecular simulation techniques, with an emphasis on tools prepared by MMTSB. Full details (including new tutorials and photographs of the conference participants) can be found at http://www.psc.edu/biomed/training/workshops/2005/Amber/index.html and documented on the MMTSB pages. A similar, but longer format, workshop was held last August (8-20), in collaboration with the Center for Theoretical Biological Physics. We had about 45 participants from more than 100 applicants for this workshop. The workshop featured the MMTSB Tool Set, CHARMM and Amber. Tutorials and lectures can be found at http://ctbp.ucsd.edu/workshopinfo_2004.html, as well as from the MMTSB web site. Finally, the first YUP training workshop was run at Georgia Tech from June 28 – July 1, 2005, and a second occurred between July 5-8. Each of these had five participants. As a result of user feedback, we have expanded the first tutorial (which teaches the use of Python) and are requiring future trainees to complete this before coming to the workshops. This enables us to shorten the workshop itself to three days. A third workshop is scheduled for August 16-18, and four people are currently enrolled. Information about the workshops and tutorials can be found at http://rumour.biology.gatech.edu/YammpWeb/, as well as on the MMTSB site.

PURDUE
UNIVERSITY

DEPARTMENT OF MEDICINAL CHEMISTRY
cbp@purdue.edu
http://bilbo.bio.purdue.edu/~postgrp

Report of the MMTSB Advisory Board
Meeting Held at the Scripps Research Institute
13 September 2004

Advisory Board Members: T. S. Baker (absent), W. Chiu, J. A. McCammon, C. B. Post (Chair), R. Scott (absent), I. Tinoco (absent)
MMTSB Investigators: C. L. Brooks III (Director), D. Case, S. Harvey, J. Johnson, Vijay Reddy, J. Skolnick, F. Tama

EXECUTIVE SUMMARY:

A primary focus of the annual Advisory Board meeting for the Research Resource for Multiscale Modeling Tools for Structural Biology (MMTSB) was the renewal application to be submitted October 2004. The report by Charles L. Brooks highlighted recent results for each of the four research objectives of the Center. All principal investigators were in attendance and presented detailed information on the past year's achievements and plans for future research. In keeping with past performance, excellent progress was made in this past year. The MMTSB Center is well positioned for submission of a renewal application. The future research plans presented to the Advisory Board are well founded and offer exciting directions for the upcoming 5-year renewal period. The Board encourages the MMTSB investigators to continue to build strong collaborations and organize workshops so that the new computational developments of the Center can be fully exploited by others.

THE REPORT:

Core research continues to be a major strength of the Center. Case and Harvey have made considerable progress in RNA modeling and developments of reduced atom representations for RNA that allow simulations of large systems on the tens of ns timescale. The generalized Born model for implicit solvent has matured and is being extended to simulations of membrane systems by Brooks. Skolnick presented significant progress toward improved accuracy of structure predictions with a new approach utilizing threading and fragment refinement, including large-scale prediction at the genome level. Both Skolnick and Brooks participated in the recent CASP6 competition. Additional computational advances were made in fitting cryo-EM data and modeling virus capsid structural dynamics. New experimental results on the time-dependence of viral capsid maturations using immobilized virus were presented by Johnson. As the time resolution of these fluorescent measurements is increased, and greater spatial resolution is achieved, Johnson aims to collaborate in computational efforts of other Center members to better understand the maturation process.

The major objectives proposed by the MMTSB Center were achieved in the current funding cycle. Over 90 publications were generated from core and collaborative research. Collaborative efforts by Center members with external investigators have increased significantly; a number of new collaborations were noted at the meeting. Software and web-based tools now available for use by the research community include the MMTSB Tool Set to support applications of CHARMM and AMBER, VIPER for analysis of icosahedral virus structure, YAMMP/YUP for RNA modeling. Dissemination was also achieved by a number of workshops organized by the MMTSB investigators. In particular, the recent molecular modeling workshop in August 2004 was the first organized by the Center in La Jolla, and was a joint effort with the NSF funded Center for Theoretical Biological Physics (CTBP). The workshop was oversubscribed and appears to have been extremely well received. The plans for a small workshop on YUP are in progress. The workshops are a critical component of the Center's dissemination efforts and should continue to receive substantial effort from all Center investigators.

The MMTSB Center is advantageously positioned. The environment of Scripps Institute and close connection with UCSD are extraordinary. In addition to the direct support provided by the RR Center, the MMTSB investigators have exploited a very strong base of institutional infrastructure. The advantages were particularly clear in the case of organizing the molecular modeling workshop and leveraging with other funded centers, CTBP and Pittsburgh Supercomputer Center.

In preparation of the renewal application it will be important to emphasize the unique strengths that distinguish the MMTSB Center. The combined expertise in theoretical chemistry, structural virology, and large scale computation are obvious strong points to justify the continued development of web-based tools such as the MMTSB toolkit, YUP and associated programs, VIPER, and future interfaces for NMFF. Documentation of literature citations, web access, and other use of these tools by the research community would provide evidence of their impact on the community. It was reported at the meeting that the Center is working to establish connections with institutions such as RCSB. Such connections clearly would be worthwhile for disseminating tools developed by MMTSB, although efforts that would duplicate services of these other institutions are seen to be less valuable. The possibility of establishing a link with a structural genomics center was also discussed. Affiliation with a structural genomics center would not only allow easy access to new data generated on a genomic scale but could also stimulate insightful discussions. Continued expansion of collaborative research efforts will be important in the next project period. Active engagement of others in collaborations, along with independent use of the methods and tools by outside investigators, would be strong testimony that the MMTSB Center provides an important service to the research community.

As new directions solidify in the next project period, Brooks may want to consider changes in the membership of the Advisory Board.